WHY INITIAL RANDOMNESS MATTERS IN PARALLEL MARKOV SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Parallel sampling aims to accelerate sequence generation by filling multiple positions per round. Yet even when every conditional distribution is known exactly, the choice of when to sample each position can limit the accuracy achievable within a fixed number of rounds. We show that, for known stationary Markov chain targets, a sampler that randomizes its first-round position set can achieve higher accuracy than any deterministic adaptive policy can achieve under the same round budget. In our sampling interface, values at the selected positions are drawn independently within each batch from exact single-coordinate conditional distributions, sampled values are never revised, and at most R rounds are used on every path. We establish this advantage for every reversible finite-state Markov chain with strictly positive transition probabilities and dependence between successive states. For each such chain, under a suitable scaling of sequence length and round budget, initial randomization yields vanishing total variation error, while the optimal deterministic adaptive error remains bounded away from zero. To quantify this advantage in forward KL divergence, we obtain matching bounds for positively correlated binary chains: over the full adaptive policy classes, randomization reduces the optimal divergence from the target to the output distribution by a factor of order 2^{R-1} in a fixed-kernel, bounded-error regime. To determine whether this randomness can be deferred, we specialize further to the symmetric binary chain in this regime and characterize how the entropy of the first-round position set controls optimal error. Attaining the optimal randomized error order requires at least $R - 1 - O(1)$ bits of this entropy, even with unrestricted later scheduling randomness and adaptation to sampled values: this initial randomness cannot be deferred.

Keywords: parallel sampling, Markov chains, randomized scheduling, adaptive policies, information theory.

1 INTRODUCTION

Parallel sampling trades sequential dependence for simultaneous draws. Even with exact single-coordinate conditional distributions (*singleton conditionals*), independently drawing values at several positions can distort their joint law. The choice of which positions to reveal, and when, is therefore central to adaptive unmasking (Lavenant & Zanella, 2025; Chen et al., 2026). At a fixed round budget, can randomizing these choices achieve higher accuracy than any deterministic adaptive policy can achieve? If so, how large is the advantage, and when must the randomness enter?

We answer these questions for known stationary Markov chain targets. A policy selects a batch of positions in each round; values at those positions are drawn independently from exact singleton conditionals given previously revealed values. Each position is revealed once, within R rounds on every path. Both deterministic and randomized policies adapt to the revealed values; randomized policies also use an independent scheduling tape. Only the sampled sequence is returned, so accuracy is measured on its marginal law after averaging over the tape.

Within this interface, the relevant object is the returned sequence law: randomization helps only if averaging over hidden schedules changes that law, not merely if it lowers an average over schedule-conditioned errors. Our results answer the three questions above with progressively sharper guarantees on more specialized chain classes.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Does scheduling randomness help? For every dependent strictly positive reversible finite-state chain, we construct a family of sequence lengths and round budgets on which a randomized sampler converges to the target in total variation, while the entire deterministic adaptive class stays a positive distance away (Section 3.1). The construction randomly shifts an equally spaced set of positions to sample in the first round, then fills the remaining intervals by exact conditional bisection.

How large is the advantage? We specialize to positive-correlation binary chains and identify the optimal forward-KL orders over both full adaptive policy classes. Write $L = 2^{R-1}$, $n = (s+1)L$, and $t = s\rho^{2L}$, where ρ is the nonconstant eigenvalue of the fixed transition kernel. In the bounded-error regime, where t stays below a fixed constant τ, under the parameter conditions stated below, the optimal deterministic and randomized errors are respectively $\Theta(t)$ and $\Theta(t/L)$. A fixed first-round grid incurs order-t error from independently sampling values that are dependent under the target distribution; randomizing its offset improves the error by a factor of order L. Theorem 1 gives the symmetric-chain bounds, and Corollary 9 covers unequal stationary frequencies.

When must the randomness enter? On the symmetric binary chain, allowing at most h nats of entropy in the first-round position set gives optimal error $\Theta(t \max\{e^{-h}, L^{-1}\})$. Thus attaining order t/L requires $\log L - O(1)$ nats initially, even with unrestricted later randomness and adaptation to sampled values. A fixed first set incurs an order-t lower bound. Theorem 2 determines both the amount of randomness needed and when it must enter.

Figure 1 compares fixed and random first batches on an eight-site symmetric chain ($\rho = 0.7$, three rounds). Rows denote separate runs; both samplers complete by exact conditional bisection. Bars show unconditional class probabilities for a single change after position k, including global flips; the dashed line is the target. Independent draws overrepresent opposite first-batch values: the fixed pair $(1, 5)$ therefore overweights $k \le 4$ and underweights $k > 4$. Uniformly choosing among $(i, i+4)$, $i = 1, \ldots, 4$, partially cancels these errors. KL uses the full output law. This is an illustration, not an optimality certificate.

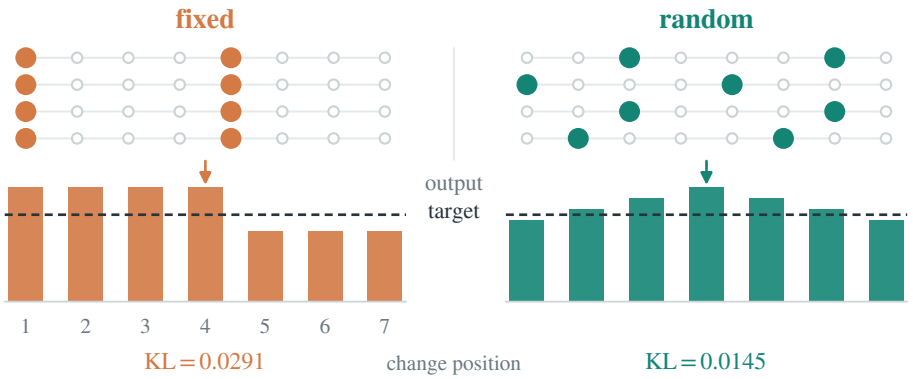

Figure 1: Random first batches reduce parallel-sampling bias.

The accuracy gap already occurs with four positions. At $n = 4$, $R = 2$, and $\rho = 1/2$ on the symmetric chain, the certified optimal forward KL drops from 0.03158 nats for deterministic adaptive policies to 0.01527 with hidden scheduling randomness (Section 6). These are full-class optima, not a comparison of two chosen schedules.

The central proof step is a converse for the returned mixture. Mixing the output laws of deterministic policies can cancel component density errors, so lower bounds on individual policy errors need not control the mixture. We instead track correlation deficits in statistics of the output sequence. Their moment bounds hold across all adaptive paths and survive averaging over the tape. Weighting these witnesses by the first-set distribution yields a lower bound in terms of the sum of its squared probabilities, and hence the entropy frontier.

After related work, Section 3 formalizes the model and three results. Sections 4 and 5 explain the accuracy gain and its first-round entropy requirement on the symmetric chain; the finite-state extension is proved in Appendix D. Section 6 supplies finite certificates of the accuracy separation.

2 Related work

Prior analyses distinguish trajectory-level factorization costs from output-marginal error. Ben-Hamu et al. (2025, Appendix A) derive this joint-to-marginal KL gap and show that it vanishes for deterministic adaptive planners; with exact singleton conditionals, their decomposition reduces to accumulated conditional total correlation. Lavenant & Zanella (2025, Proposition 2) bound marginal KL for randomized, value-dependent planners by learning and factorization errors. Wen et al. (2026, Theorem 1 and Remark 2) formulate the deterministic adaptive total-correlation identity and leave arbitrary randomized-policy converses open in their setting; their target-averaged serial depth differs from our pathwise round cap. DUEL develops the likelihood consequences of deterministic adaptive and latent stochastic unmasking (Turok et al., 2026).

Chen et al. (2026, Theorem 16) bound randomized output-mixture KL; their Theorems 24–25 give output-marginal TV lower bounds for randomized adaptive conditional-oracle algorithms over unknown target families. Our converses fix a *known* target and cover all adaptive policies with exact-singleton, independent-batch, reveal-once access and a pathwise round cap. Average component KL lower bounds cannot establish such output-mixture converses. Li & Cai (2026) analyze mask-averaged component KL; Zhao & Cai (2026) randomize batch sizes with KL guarantees governed by total or dual total correlation. Randomized unmasking is therefore not new: we establish its advantage over *every* deterministic adaptive policy at the same round cap, and the necessity of initial randomness. Wainwright (2026) optimize schedules through estimable unmasking-growth certificates, rather than full-class output-law converses. Dependent-variable martingale concentration (Kontorovich & Ramanan, 2008) provides methodological background; we derive concentration under the adaptive output law, not just the target.

Our construction combines nonadjacent coarse-to-fine reveals, as in dilated scheduling (Luxembourg et al., 2025), exact Markov-separator bisection (Bansal et al., 2026, Lemma 4), and order-probability mixtures (Uria et al., 2014, Section 4). The novelty is their hidden-phase mixture's separation from the full deterministic adaptive class, not these ingredients individually.

Speculative verification or rejection (Anari et al., 2024; 2025) and Gibbs scan-order comparisons (He et al., 2016) fall outside our irrevocable independent reveals. Feng et al. (2025) distinguish token-level and sequence-error guarantees, including hidden-Markov targets; these metrics differ from our output-law KL and TV. Sampler-centric evaluations using exact Markov-model posterior oracles also isolate sampling error from denoiser error (Tang et al., 2026); they study specified samplers and transition-level diagnostics, rather than full-class output-law converses.

PUNT selects batches through conditional-independence tests and confidence (Azangulov et al., 2025). Entropy-bounded unmasking (Ben-Hamu et al., 2025) and confidence-based decoding (Cai & Li, 2026) use predictive entropy to control batching or relate iteration complexity to data entropy. These differ from the entropy of our *first-round position set*. Variable-length resolvability studies the random input length needed to approximate a law (Yagi & Han, 2017); exact common information studies shared randomness for distributed generation with local randomness (Kumar et al., 2014). Neither imposes roundwise reveals. We charge only first-set entropy, leaving oracle draws and later randomness free, and ask whether later randomness can replace initial randomness.

3 Setup and main results

We compare deterministic adaptive schedules with hidden randomized schedules under the same pathwise round cap. The singleton oracle is exact: approximation arises only from drawing conditionally dependent positions independently within a batch.

Let P_n be the stationary path law of a known strictly positive Markov kernel on a finite alphabet. A history h records the revealed positions A, their values x_A, and the scheduling transcript and state. The policy selects a nonempty batch of positions $B \subseteq [n] \setminus A$, which need not be contiguous. Given h and B, a value at each position $j \in B$ is drawn independently from the exact singleton conditional $P_n(X_j \mid X_A = x_A)$. The new values become available only after the whole batch has been sampled. Draws are irrevocable, and every position is revealed exactly once within R rounds.

A deterministic policy chooses B as a function of h. A randomized policy chooses it as a function of (I, h), where I is an independent scheduling tape encoding all initial and later scheduling randomness.

Its law may depend on the known target and round budget and has unrestricted entropy. Both classes are adaptive: they may use the full history, including sampled values. "Deterministic" describes the scheduling rule, not the sampled values. Each batch is chosen before its values are drawn, and the oracle's private sampling coins remain inaccessible to the scheduler. The round cap holds for every tape in the support and every positive-probability sequence of oracle outcomes.

Write $\mathcal{D}_R$ and $\mathcal{R}_R$ for these deterministic and randomized adaptive classes. The returned sequence has law Q_σ after averaging over the scheduling tape. We measure total-sequence forward KL $D_{\mathrm{KL}}(P_n\|Q_\sigma)$, using natural logarithms, and total variation.

The results below distinguish *scope*, *size*, and *timing*: TV establishes a separation for general reversible chains; matching KL bounds quantify its size for binary chains; first-set entropy measures the initial randomness needed on the symmetric chain.

3.1 Distributional separation for finite-state chains

Fix a strictly positive reversible kernel K with stationary law π. Assume it is dependent, meaning $K(x,y) \not\equiv \pi(y)$, and write its nonconstant spectral radius as $\rho = \max_{j\geq 2} |\lambda_j(K)| \in (0,1)$. Set $L = 2^{R-1}$, $n = (s+1)L$, and $t = s\rho^{2L}$ for integers $R, s \geq 1$. Choose a grid offset i uniformly from $\{1,\dots,L\}$ and sample values at positions $i, i+L, \dots, i+sL$ in the first round. Exact conditional completion of the remaining positions defines the output law $\overline{Q}$. We also refer to the grid offset i as its *phase*. Theorem 10 gives deterministic optimal KL of order t and randomized KL $O_{K,\tau}(t/L)$ when $0 < t \leq \tau$ and $\rho^L/\min_x \pi(x) \leq 1/2$. The separation also holds in total variation: Corollary 11 establishes

$$\inf_{\sigma\in\mathcal{D}_R} \mathrm{TV}(P_n, Q_\sigma) \geq 1 - e^{-a_K t}, \qquad \mathrm{TV}(P_n, \overline{Q}) = O_{K,\tau}\left(\sqrt{t/L}\right), \tag{1}$$

where $a_K > 0$. Choosing $s = \lfloor t_0/\rho^{2L} \rfloor$ for fixed $t_0 > 0$ and letting $L \to \infty$ makes the randomized error vanish while the optimal deterministic error stays positive. The proof combines adaptive Hellinger-affinity bounds with phase-mixture concentration (Appendices D.2 and D.3).

3.2 The sharp binary accuracy frontier

For binary chains with positive correlation, the randomized rate t/L also has a matching converse, including unequal stationary frequencies (Corollary 9). We develop the accuracy and first-set entropy frontiers on the symmetric chain: fix $\rho \in (0,1)$, take a uniform initial spin on $\{-1,+1\}$, and use the transition kernel

$$K(x,y) = \frac{1+\rho xy}{2}. \tag{2}$$

The next theorem states the resulting accuracy frontier. Its constructive upper bound and both converses are unpacked in Section 4.

Theorem 1 (Optimal randomized and deterministic orders)**.** *Fix $0 < \rho < 1$ and $0 < \tau < \infty$. Let $R, s \geq 1$ be integers, $L = 2^{R-1}$, $n = (s+1)L$, and $t = s\rho^{2L}$. For the stationary chain* (2) *and the complete policy classes of Section 3, with $\rho^L \leq 1/4$ and $0 < t \leq \tau$, the optimal marginal error $\inf_{\sigma\in\mathcal{C}} D_{\mathrm{KL}}(P_n\|Q_\sigma)$ satisfies:*

Adaptive policy class $\mathcal{C}$	Lower bound	Achievable upper bound
Deterministic $\mathcal{D}_R$	$d_\rho t$	$2t$ (fixed phase)
Hidden randomized $\mathcal{R}_R$	$c_\rho t/L$	$C_{\rho,\tau} t/L$ (random phase)

Here $d_\rho, c_\rho > 0$ depend only on ρ, and $C_{\rho,\tau} < \infty$ depends only on ρ, τ; explicit bounds appear in Appendices B.1 and B.3. The randomized upper bound follows from Theorem 5. The randomized lower bound is Theorem 8. Any fixed phase is deterministic and has error $s\phi(\rho^L) \leq 2t$ by (7)*; its converse is* (12)*. Thus the two optimal errors differ by a factor of order L. These bounds match in order for fixed kernel and error-regime parameters; the bounds are not intended as tight finite-instance estimates.*

3.3 THE FIRST-SET ENTROPY FRONTIER

Let $H_0(\sigma) = H(B_0)$ be the Shannon entropy, in nats, of the first-round position set B_0 (the *first set*), and put $\mathcal{E}_R(h) = \{\sigma \in \mathcal{R}_R : H_0(\sigma) \le h\}$. This counts neither randomness used to sample values nor later scheduling randomness. In particular, $H_0 = 0$ permits arbitrary randomized adaptive continuations after a fixed first set.

Theorem 2 (First-set entropy frontier)**.** *Fix $0 < \rho < 1$ and $0 < \tau < \infty$. For the stationary symmetric chain* (2)*, let $R, s \ge 1$ be integers, $L = 2^{R-1}$, $n = (s+1)L$, and $t = s\rho^{2L}$. If $\rho^L \le 1/4$ and $0 < t \le \tau$, then for every $h \ge 0$,*

$$\inf_{\sigma \in \mathcal{E}_R(h)} D_{\mathrm{KL}}(P_n \| Q_\sigma) = \Theta_{\rho,\tau}\big(t \max\{e^{-h}, L^{-1}\}\big). \tag{3}$$

Uniformly choosing $\min\{L, \lfloor e^h \rfloor\}$ distinct grid phases attains this order. Zero first-set entropy leaves optimal error of order t, even with unrestricted later scheduling randomness.

Under the same assumptions, restricting the policy to at most M possible first sets gives optimal error $\Theta_{\rho,\tau}(t/\min\{M, L\})$. More precisely, for every fixed accuracy factor $C > 0$, reaching $D_{\mathrm{KL}}(P_n\|Q) \le Ct/L$ requires $H_0(\sigma) \ge \log L - O_{\rho,\tau,C}(1)$ nats initially; unlimited later randomness cannot replace that requirement. Section 5 explains the collision-probability lower bound; Appendix C proves both frontiers.

We now develop the two binary frontiers: Section 4 explains the accuracy orders, and Section 5 isolates the first-set obstruction. The broader finite-state result is proved in Appendix D.

4 THE VALUE OF SCHEDULING RANDOMNESS

The main theorem separates an achievable phase mixture from converses over complete adaptive policy classes. We now unpack those two sides: first the cancellation mechanism, then the lower bounds that show its order is optimal. Full upper-bound proofs and explicit constants appear in Appendices B.2 and B.1.

4.1 ACHIEVING THE RANDOMIZED RATE

Write $u = \rho^L$, $v = u^2$, and $t = sv$.

For a grid offset $i \in \{1, \dots, L\}$, independently sample values at the $s+1$ positions $i, i+L, \dots, i+sL$ from their uniform target marginals in round one. Each remaining interval, including the boundary tails, has at most $L-1$ sites. In every later round, select a middle position (a *separator*) in each nonempty interval and sample its value from its exact singleton conditional given previously revealed values. The intervals are conditionally independent, so this completes the target conditional law given the first-round values in at most $R-1$ additional rounds. Denote the resulting component law by Q_i. Choosing the offset uniformly gives $\overline{Q} = L^{-1}\sum_i Q_i$, using $R-1$ external scheduling bits within the same R-round cap (Bansal et al., 2026). Each run chooses one offset and returns one sequence, whose law is the mixture $\overline{Q}$.

The gain comes from cancellation, not from making every fixed phase accurate. It can be quantified in two complementary ways: a transcript-level cost for a fixed schedule and an output-marginal cost after the scheduling tape is hidden. For a fixed deterministic policy, the history-conditioned batch cost is

$$C(h) = \mathrm{TC}(X_B \mid h) = \sum_{j \in B} H(X_j \mid h) - H(X_B \mid h), \qquad D_{\mathrm{KL}}(P_n\|Q_\sigma) = \mathbb{E}_{P_n} \sum_r C(h_r). \tag{4}$$

The expectation follows histories obtained by replaying the policy on a target draw. For a discrete independent scheduling tape with prior μ, let Q_i be the output law of the deterministic policy obtained by fixing the tape to i. For the mixture $\overline{Q} = \sum_i \mu_i Q_i$, the ordinary KL chain rule instead gives

$$\sum_i \mu_i D_{\mathrm{KL}}(P_n\|Q_i) = D_{\mathrm{KL}}(P_n\|\overline{Q}) + \mathbb{E}_{X\sim P_n} D_{\mathrm{KL}}\big(\mu\|\nu(\cdot \mid X)\big), \quad \nu(i \mid x) = \frac{\mu_i Q_i(x)}{\overline{Q}(x)}. \tag{5}$$

The posterior term is exactly the reduction in KL obtained by hiding the tape. Signed density errors tied to different scheduling decisions can cancel in the returned mixture. Theorem 1 characterizes the attainable reduction and the remaining marginal error on the symmetric chain.

Under P_n, the edge spins $Z_e = X_e X_{e+1}$ are independent with mean ρ. Write

$$Y_{i,j} = \prod_{e=i+jL}^{i+(j+1)L-1} Z_e, \qquad T_i = \frac{Q_i}{P_n} = \prod_{j=0}^{s-1} (1 + uY_{i,j})^{-1}, \qquad T = \frac{\overline{Q}}{P_n} = \frac{1}{L}\sum_{i=1}^{L} T_i. \tag{6}$$

Conditional completion is exact, so the density ratio only reflects replacing the joint target law on the first-round positions by independent uniform spins. The windows within one phase are disjoint. Hence every component has exactly

$$D_{\mathrm{KL}}(P_n \| Q_i) = s\phi(u), \qquad \phi(u) = \tfrac{1}{2}\log(1-u^2) + u\,\mathrm{atanh}(u). \tag{7}$$

Since $\phi(u) = u^2/2 + O(u^4)$, the component cost and posterior term both approach $t_0/2$ in Corollary 3, while the mixture error vanishes.

Two estimates turn this cancellation into an upper bound. First, averaging phases concentrates the density ratio. Each $T_i - 1$ has mean zero under P_n and contains both excess and missing probability mass. The cross-phase calculation shows that their aggregate overlap is controlled by geometric spatial decay. In the regime $u \le 1/4$ and $0 < t \le \tau$,

$$V := \mathrm{Var}_{P_n}(T) = \chi^2(\overline{Q} \| P_n) = O_{\rho,\tau}(t/L).$$

Appendices B.2.2–B.2.3 give the transfer calculation and exact phase-pair sum.

Second, forward KL also requires control of the lower tail: small T means that the mixture assigns too little mass to a target sequence. Set $A = -L^{-1}\sum_i \log T_i$. AM–GM gives $-\log T \le A$, while covariance counting gives $\mathbb{E}A = s\phi(u) \le 2t$ and $\mathrm{Var}(A) = O_\rho(t/L)$. Splitting at $T = 1/2$ and centering A on the lower-tail event yields

$$D_{\mathrm{KL}}(P_n \| \overline{Q}) \le (2 + 4\mathbb{E}A)V + 2\sqrt{\mathrm{Var}(A)V} = O_{\rho,\tau}(t/L).$$

Thus density concentration and log-cost concentration play different roles; both are needed for this forward-KL bound. Appendix B.2.4 gives the lower-tail argument and Appendix B.1 the constants.

4.2 The deterministic lower bound

For deterministic policies, the total-correlation identity (4) turns unavoidable within-batch dependence into an output-KL lower bound.

Partition the chain into disjoint cells of $2L = 2^R$ sites. Every labeling of one such cell with R reveal-round labels contains two sites at distance at most L with equal labels and no smaller label between them. This short-witness fact follows by induction on R.

At a deterministic history, the event that the policy reached that history is exactly the cylinder specifying its revealed values. For an eligible witness pair at distance $d \le L$, its conditional law therefore has the form

$$p(x,y) \;\propto\; (1+\alpha x)(1+\rho^d xy)(1+\beta y), \qquad |\alpha|, |\beta| \le \rho.$$

Its covariance is $\rho^d(1-\alpha^2)(1-\beta^2)/(1+\alpha\beta\rho^d)^2$. Pinsker's inequality gives a mutual-information lower bound uniform over all revealed values. Charge each cell the first time a qualifying pair is selected, before sampling that round. The pairs charged at a node are disjoint, so entropy subadditivity bounds the sum of their mutual informations by the batch total correlation. Every completed path charges every cell once. Applying (4) gives $D_{\mathrm{KL}}(P_n \| Q_\sigma) \ge d_\rho t$, with $d_\rho > 0$ depending only on ρ. The charging uses only the history available before each batch is drawn.

4.3 The randomized lower bound

The deterministic argument does not extend by averaging over the scheduling tape: the posterior term in (5) can absorb the component costs. The following output-only argument controls the returned mixture directly.

We therefore change from transcript costs to an output-only statistic. The converse uses a statistic of the returned sequence, so its bound survives marginalizing any independent scheduling tape. We outline the proof of $D_{\mathrm{KL}}(P_n\|Q_\sigma) \geq c_\rho t/L$ for every $\sigma \in \mathcal{R}_R$; Appendix B.3 supplies the details.

Define the weighted correlation statistic and its target conditional mean

$$F_L(x) = \sum_{d=1}^{L} \rho^{L-d} \sum_{i=1}^{n-d} x_i x_{i+d}, \qquad M(A, x_A) = \mathbb{E}_{P_n}[F_L \mid X_A = x_A], \quad \mu = \mathbb{E}_{P_n} F_L.$$

Write $Q = Q_\sigma$. Let A_r be the revealed set after r rounds, and let B_r be the batch selected for round $r+1$, with $0 \leq r < R$. Completed paths are padded by empty batches.

The weights turn each distance-d covariance of order ρ^d into order ρ^L. Order n/L unavoidable witnesses and an order-n moment scale then give error of order $n\rho^{2L}/L^2 \asymp t/L$.

First, product updates lose conditional correlation. Conditional pair moments have nonnegative quadratic boundary-spin coefficients. Independent singleton updates therefore lose only nonnegative covariances. If $M_r = M(A_r, X_{A_r})$ and the filtration includes the independent scheduling tape, its predictable deficit is $D_r = M_r - \mathbb{E}_{\widetilde{Q}}[M_{r+1} \mid \mathcal{F}_r] \geq 0$, where $\widetilde{Q}$ is the joint tape/output law.

Second, every completed path incurs a nonzero cumulative deficit. An online charging argument forces one same-round pair at distance at most L in each disjoint $2L$-cell, with no previously revealed site between the pair. Its covariance gives weighted deficit at least $\gamma_\rho \rho^L$. Thus $\sum_r D_r \geq \delta := \lfloor n/(2L) \rfloor \gamma_\rho \rho^L$ on every path. Each charge is determined from the current revealed history.

Finally, a uniform output-law moment bound survives mixing. Coupling the target conditional laws with one revealed spin fixed to $+1$ and -1, while keeping all other revealed values unchanged, bounds the resulting change in M by C_ρ. Conditional Hoeffding bounds yield a compensated exponential supermartingale. Its terminal cost is $C_\rho^2 n$, since every position is revealed exactly once, even under adaptive batch choices. With $\eta_r = M_{r+1} - \mathbb{E}_{\widetilde{Q}}[M_{r+1} \mid \mathcal{F}_r]$ and $F_L - \mu = \sum_r (\eta_r - D_r)$, for $\lambda \geq 0$ the *marginal* law obeys

$$\mathbb{E}_Q e^{\lambda(F_L - \mu)} \leq e^{-\lambda\delta + \lambda^2 C_\rho^2 n/8}, \qquad D_{\mathrm{KL}}(P_n\|Q) \geq \lambda\delta - \lambda^2 C_\rho^2 n/8.$$

The KL variational inequality gives the second bound, since $\mathbb{E}_{P_n}(F_L - \mu) = 0$. Optimizing λ gives $D_{\mathrm{KL}}(P_n\|Q_\sigma) \geq c_\rho t/L$ for every admissible randomized policy; Appendix B.3 gives the explicit constant.

Together with the phase construction, these deterministic and randomized converses establish the sharp orders in Theorem 1. The fixed-total-error consequence is:

Corollary 3 (Fixed total-error scale)**.** *Fix $0 < \rho < 1$ and $t_0 > 0$. Along dyadic $L \to \infty$, let $s = \lfloor t_0/\rho^{2L} \rfloor$, $n = (s+1)L$, and $R = 1 + \log_2 L$. The random-offset error is $O_{\rho,t_0}(1/L)$, whereas the infimum error over $\mathcal{D}_R$ has a positive lower limit.*

Along this family, $L = \Theta_\rho(\log n)$ and $R = \log_2 \log n + O_\rho(1)$, so the accuracy improvement is of order $\log n$ at the same round cap. Appendix B.2.9 also treats growing t.

5 The timing of scheduling randomness

The preceding section establishes the accuracy gain and its optimal order. It remains to locate where the scheduling randomness must be spent. The random-offset construction chooses its first set before observing any values. Theorem 2 shows that this timing is essential: a fixed first set leaves an order-t lower bound even with unrestricted randomized adaptive continuations. To quantify the required initial spread, we weight the output witness by the distribution of first sets.

5.1 A lower bound from first-set collision probability

Write $w_b = \Pr(B_0 = b)$ and $c_2 = \sum_b w_b^2$, the collision probability of two independent first sets. We construct a single output statistic determined by this distribution before any continuation is observed.

The key is to cover every continuation after a first set has been selected. In each $2L$-cell, either the first set contains a pair at distance at most L, or the remaining rounds force a same-round pair at

distance at most $L/2$ with no earlier revealed site between its endpoints. Later randomization can change which pair witnesses the loss, but cannot avoid all such pairs. When $R = 1$, every site is in the first set, so only the first case occurs.

We therefore fix a collection E of pairs before any continuation is observed. For each possible first set b, designate one first-round pair of distance at most L in each cell where one exists; also include every within-cell pair of distance at most $L/2$ to cover the later witnesses. A designated pair receives the largest w_b among first sets designating it; each short pair receives $w_* := \max_b w_b$, taking the larger weight when both rules apply. Writing these weights as v_{jk}, define

$$F(x) = \sum_{(j,k)\in E} \rho^{L-(k-j)} v_{jk} x_j x_k.$$

Unlike F_L, this statistic depends on the first-set distribution. For every path starting with b, each cell supplies a witness with $v_{jk} \geq w_b$. The distance weight turns its covariance loss into order $\rho^L v_{jk}$, so the cumulative deficit is at least $d_0 w_b$, where d_0 is of order $(n/L)\rho^L$.

The extra pairs have distance at most $L/2$ and hence strongly discounted weights $\rho^{L-(k-j)}$. The cover's squared-weight mass is $O_\rho((n/L)c_2)$. This also bounds the total oscillation cost S, the sum of squared coordinate-change bounds for the target conditional means of F. The compensated exponential-moment argument from the randomized converse in Section 4.3 then gives, for the actual output marginal Q,

$$\mathbb{E}_Q e^{\lambda(F-\mathbb{E}_{P_n}F)} \leq e^{\lambda^2 S/8} \sum_b w_b e^{-\lambda d_0 w_b}, \qquad \lambda \geq 0.$$

The same statistic and moment control work for every continuation, so hiding the later scheduling choices preserves this bound. Controlling the remaining fluctuation of w_b and optimizing the KL variational bound gives

$$D_{\mathrm{KL}}(P_n \| Q) \geq \frac{\kappa_\rho t}{1 + \kappa_\rho t} \sum_b w_b^2, \qquad \kappa_\rho > 0. \tag{8}$$

Appendix C gives the cover and moment estimates in full.

5.2 The entropy–accuracy tradeoff

Two examples explain the collision bound when $t \leq \tau$. A fixed first set has $c_2 = 1$, leaving an order-t lower bound regardless of later randomness. A uniform choice among q distinct first sets has $c_2 = 1/q$, giving a lower bound of order t/q. Uniformly mixing $q \leq L$ grid phases attains this order. The lower bound applies to arbitrary first sets, and grid phases supply the matching construction. The universal t/L lower bound prevents further order improvement once q reaches L.

More generally, $c_2 \geq e^{-H(B_0)}$. Combining the collision bound with the universal t/L converse proves the lower half of Theorem 2. For the upper half, mix $q = \min\{L, \lfloor e^h \rfloor\}$ distinct phases uniformly. This uses entropy $\log q \leq h$, and the phase-overlap and lower-tail estimates give $O(t/q)$. The achieving policy spends all its scheduling randomness in the first set; the converse permits unrestricted later randomness and adaptation to sampled values.

The collision bound is sharper for a specified nonuniform first-set law; the entropy theorem instead optimizes over all policies with $H(B_0) \leq h$. At $H(B_0) = 0$, the order-t lower bound covers all randomized adaptive continuations after a fixed first set, strengthening the deterministic-policy converse.

6 Finite-policy certificates

The asymptotic frontiers establish the orders of the effect. Finite-policy optimization now provides certified separations at small sequence lengths (Table 1). These comparisons illustrate the accuracy gain; the first-set entropy requirement is established analytically in Section 5.

At $R = 2$, the $2^n - 1$ nonempty first-set choices determine all deterministic output laws; their convex hull gives all hidden-randomized laws. Rational certificates and interval arithmetic certify the optimal

KL values to at least six significant digits (Appendix E.1). At $R = 3$, a Bellman recurrence computes the full adaptive deterministic optimum, which we compare with the uniform four-phase mixture (Appendix E.2). Its divergence D_{mix} is a feasible value, not a certified randomized optimum. Table 1 shows the $n = 12$ three-round cases; Table 3 lists all six settings.

Table 1: Certified forward KL (nats) at two and three rounds.

R	n	ρ	D^*_{det}	D^*_{rand}
2	4	0.50	0.0315839	0.0152664
2	6	0.50	0.0631679	0.0304827
2	8	0.50	0.0947518	0.0467582
R	n	ρ	D^*_{det}	D_{mix}
3	12	0.25	1.52588×10^{-5}	4.25530×10^{-6}
3	12	0.50	0.00390880	0.00150589
3	12	0.75	0.0819161	0.0629505

Proposition 4 (Certified finite strict separation)**.** *For the stationary symmetric binary chain with $n = 4$, $\rho = 1/2$, and round cap $R = 2$, the complete adaptive deterministic and hidden-randomized optima satisfy $D^*_{\mathrm{rand}} < D^*_{\mathrm{det}}$, certified by*

$$D^*_{\mathrm{det}} \in [0.031583942401, 0.031583942403], \qquad D^*_{\mathrm{rand}} \in [0.015266397360, 0.015266397362].$$

The same four-site target also has a strict TV separation (Appendix E.1.1). At $R = 3$, the four-phase mixture improves on the full deterministic optimum at all six settings $n \in \{8, 12\}$, $\rho \in \{1/4, 1/2, 3/4\}$, by factors 1.30–3.64. These finite comparisons complement the asymptotic accuracy frontiers.

7 Discussion

Hidden scheduling randomness yields a distributional separation across dependent strictly positive reversible finite-state chains. For positive-correlation binary chains, it improves the optimal KL order by a factor $L = 2^{R-1}$ at a fixed round budget. On the symmetric chain, attaining this order asymptotically requires $\log L - O(1)$ nats of first-set entropy, even with unrestricted later randomness and adaptation to sampled values. Along the fixed-t family, $L = \Theta_\rho(\log n)$ and $R = \log_2 \log n + O_\rho(1)$, so the accuracy gain is of order $\log n$. An additional deterministic round asymptotically reduces error below the optimal R-round randomized order (Appendix B.2.9). Thus first-set entropy and round count are distinct resources, and the fixed-budget separation does not imply an unbounded round saving.

Our analysis assumes a known stationary Markov target and exact singleton conditionals. Matching randomized lower bounds for the full reversible finite-state class remain open: our matching results cover positive-correlation binary chains and the product extensions in Section D.1. Sharp accuracy and first-set entropy frontiers for learned or inexact conditionals are also not established here.

Reproducibility statement

Appendices B–D provide complete proofs. Appendix E gives the finite policy-class certificates, deterministic dynamic program, and rational verification principles. The accompanying artifact documents independent checks, exact certificate replay, numerical settings, and execution commands.

AI use statement

Writing and polishing. Generative AI tools assisted with revising the paper's organization, paragraph transitions, terminology, and figure captions.

Literature retrieval and discovery. For Section 2, AI tools helped identify related work and compare its assumptions, sampling interfaces, and guarantees with ours.

Research ideation and execution. We proposed the initial research question and refined the research focus and scope through discussions with AI tools. AI tools also assisted in formulating and refining mathematical claims, their assumptions, and their scope of applicability. AI tools helped design the finite-instance verification methodology, implement the computations reported in Section 6, develop and run the certificate-verification scripts described in Appendix E, interpret the computational results, and prepare figures.

Drafting manuscript sections. AI tools helped draft sections of the manuscript and revise explanations of the setup, main results, and proof mechanisms in Sections 3–5.

Proving mathematical claims. For the arguments in Sections 4 and 5 and the proofs in Appendices B–D, AI tools assisted in developing proof strategies and derivations, drafting and revising proofs, checking algebraic identities, and scrutinizing the randomized output-marginal and first-set entropy arguments, including reviews by separate AI agents.

REFERENCES

Nima Anari, Ruiquan Gao, and Aviad Rubinstein. Parallel sampling via counting. arXiv preprint arXiv:2408.09442, 2024. URL https://arxiv.org/abs/2408.09442v1. Version 1, August 18, 2024.

Nima Anari, Carlo Baronio, CJ Chen, Alireza Haqi, Frederic Koehler, Anqi Li, and Thuy-Duong Vuong. Parallel sampling via autospeculation. arXiv preprint arXiv:2511.07869, 2025. URL https://arxiv.org/abs/2511.07869v1. Version 1, November 11, 2025.

Iskander Azangulov, Teodora Pandeva, Niranjani Prasad, Javier Zazo, and Sushrut Karmalkar. Parallel sampling from masked diffusion models via conditional independence testing. arXiv preprint arXiv:2510.21961, 2025. URL https://arxiv.org/abs/2510.21961v1. Version 1.

Vansh Bansal, Cho Cholyeon, Syamantak Kumar, Sujay Sanghavi, and Purnamrita Sarkar. Understanding parallel samplers in masked diffusion via random walks on graphs. arXiv preprint arXiv:2606.22976, 2026. URL https://arxiv.org/abs/2606.22976v1. Version 1, June 22, 2026.

Heli Ben-Hamu, Itai Gat, Daniel Severo, Niklas Nolte, and Brian Karrer. Accelerated sampling from masked diffusion models via entropy bounded unmasking. arXiv preprint arXiv:2505.24857, 2025. URL https://arxiv.org/abs/2505.24857v1. Version 1, May 30, 2025.

Changxiao Cai and Gen Li. Confidence-based decoding is provably efficient for diffusion language models. arXiv preprint arXiv:2603.22248, 2026. URL https://arxiv.org/abs/2603.22248v1. Version 1.

Sitan Chen, Kevin Cong, and Jerry Li. Optimal inference schedules for masked diffusion models. In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 1279–1311. PMLR, 2026. URL https://proceedings.mlr.press/v336/chen26e.html.

Guhao Feng, Yihan Geng, Jian Guan, Wei Wu, Liwei Wang, and Di He. Theoretical benefit and limitation of diffusion language model. arXiv preprint arXiv:2502.09622, 2025. URL https://arxiv.org/abs/2502.09622v2. Version 2, June 9, 2025.

Bryan He, Christopher De Sa, Ioannis Mitliagkas, and Christopher Ré. Scan order in Gibbs sampling: Models in which it matters and bounds on how much. arXiv preprint arXiv:1606.03432, 2016. URL https://arxiv.org/abs/1606.03432v1. Version 1, June 10, 2016.

Leonid (Aryeh) Kontorovich and Kavita Ramanan. Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability*, 36(6):2126–2158, 2008. doi: 10.1214/07-AOP384. URL https://arxiv.org/abs/math/0609835v2. Author reprint read: arXiv:math/0609835v2, January 19, 2009.

Andreas Krause and Carlos Guestrin. Optimal nonmyopic value of information in graphical models—efficient algorithms and theoretical limits. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, 2005. URL https://www.cs.cmu.edu/~guestrin/Publications/IJCAI2005/ijcai2005.pdf.

Gowtham Ramani Kumar, Cheuk Ting Li, and Abbas El Gamal. Exact common information. *arXiv preprint arXiv:1402.0062*, 2014. Version 1, February 1, 2014.

Hugo Lavenant and Giacomo Zanella. Error bounds and optimal schedules for masked diffusions with factorized approximations. arXiv preprint arXiv:2510.25544, 2025. URL https://arxiv.org/abs/2510.25544v2. Version 2, December 17, 2025.

Gen Li and Changxiao Cai. Breaking AR's sampling bottleneck: Provable acceleration via diffusion language models. arXiv preprint arXiv:2505.21400, 2026. URL https://arxiv.org/abs/2505.21400v2. Version 2, January 8, 2026; first posted in 2025.

Omer Luxembourg, Haim Permuter, and Eliya Nachmani. Plan for speed: Dilated scheduling for masked diffusion language models. arXiv preprint arXiv:2506.19037, 2025. URL https://arxiv.org/abs/2506.19037v5. Version 5, May 25, 2026.

Luhan Tang, Longxuan Yu, Shaorong Zhang, and Greg Ver Steeg. Is your diffusion sampler actually correct? a sampler-centric evaluation of discrete diffusion language models. arXiv preprint arXiv:2602.19619, 2026. URL https://arxiv.org/abs/2602.19619v2. Version 2, May 27, 2026.

Gilad Turok, Chris De Sa, and Volodymyr Kuleshov. DUEL: Exact likelihood for masked diffusion via deterministic unmasking. arXiv preprint arXiv:2603.01367, 2026. URL https://arxiv.org/abs/2603.01367v2. Version 2, March 10, 2026.

Benigno Uria, Iain Murray, and Hugo Larochelle. A deep and tractable density estimator. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, 2014. URL https://proceedings.mlr.press/v32/uria14.pdf.

Martin J. Wainwright. The data geometry of masking diffusion: Certified-optimal schedules via unmasking growth complexity. arXiv preprint arXiv:2608.13520, 2026. URL https://arxiv.org/abs/2608.13520v1. Version 1, August 13, 2026.

Chuling Wen, Weijie Liang, and Jian Lu. Conditional total correlation and the serial depth of adaptive parallel sampling. arXiv preprint arXiv:2608.25505, 2026. URL https://arxiv.org/abs/2608.25505v1. Version 1, August 26, 2026.

Hideki Yagi and Te Sun Han. Variable-length resolvability for general sources and channels. *arXiv preprint arXiv:1701.08712*, 2017. Version 1, January 30, 2017.

Yunxiao Zhao and Changxiao Cai. Adaptation to intrinsic dependence in diffusion language models. arXiv preprint arXiv:2602.20126, 2026. URL https://arxiv.org/abs/2602.20126v1. Version 1, February 23, 2026.

A PROOF ROADMAP AND CONVENTIONS

The proof order starts with the symmetric binary chain, where the mechanism is explicit, then treats the timing of randomness and the finite-state extension. Throughout, we use the interface of Section 3: positions are revealed once, batches use independent exact singleton conditionals, and every path finishes within the round cap. The scheduling tape is independent of the oracle's private draws and is marginalized out of the returned law. Targets are stationary; logarithms and entropies are in nats.

Accuracy: construction, then converse (Appendix B). For the symmetric-chain frontier, read Appendix B.2 for the hidden-phase upper bound and deterministic lower bound, then Appendix B.3 for the randomized marginal lower bound. The upper bound follows the chain: exact separator completion, cross-phase moments, density variance, and forward KL. The deterministic converse uses three reusable ingredients: the target-history identity (Appendix B.2.5), the reveal-label witness (Lemma 6), and online cell charging (Appendix B.2.8). These ingredients do not depend on a binary alphabet; only the conditional pair-information bound does. The randomized converse instead controls a statistic of the returned sequence, since componentwise KL lower bounds do not control a mixture. Appendix B.4 extends this converse to unequal binary state frequencies; its upper and deterministic bounds use Theorem 10, proved later.

Timing: first-set probabilities, then entropy (Appendix C). The entropy frontier reweights the symmetric-chain witnesses by first-set probabilities. Collision probability gives the entropy-sensitive lower bound; Appendix B.3 supplies the universal t/L floor. Mixing a controlled number of phases gives the matching upper bound. This argument permits unrestricted later scheduling randomness.

Breadth: spectral estimates, then output discrimination (Appendix D). For the general finite-state result, read Theorem 10, its quantitative proof in Appendix D.2, and the affinity argument in Appendix D.3. The new work is spectral transfer and boundary control; the history and charging arguments are reused rather than reproved. The TV lower bound requires its own affinity contraction, not an inference from a KL lower bound. Negative and repeated spectral modes are allowed. Matching randomized converses are established only for the stated binary and independent-product cases, not all finite-state chains.

Verification: distinguish the certified quantities (Appendix E). Appendix E.1 certifies the complete two-round policy classes. Appendix E.2 optimizes the full deterministic adaptive class at larger round budgets; its three-round random mixture is only a feasible comparator. Appendix E.3 evaluates the analytic bounds with rational arithmetic. These certificates complement, rather than replace, the proofs; execution details remain in the research artifact.

B BINARY-CHAIN ACCURACY FRONTIERS

B.1 EXPLICIT CONSTRUCTION AND DETERMINISTIC BOUNDS

The following constants quantify the symmetric-chain construction and full deterministic converse in Theorem 1. Appendix B.2 proves these bounds; the randomized lower bound and its constants are stated together with their proof in Appendix B.3.

Theorem 5 (Hidden-phase separation). *Fix $0 < \rho < 1$ and integers $R, s \geq 1$. For the stationary kernel* (2)*, let $L = 2^{R-1}$, $n = (s+1)L$, and put $u = \rho^L \leq 1/4$, $v = u^2$, and $t = sv$. Define*

$$B_\rho = \frac{1+\rho^2}{1-\rho^2}, \qquad d_\rho = \frac{(1-\rho^2)^4}{4(1+\rho^3)^4}, \tag{9}$$

$$V_* = \frac{e^{2t} - 1 + 20te^{10t}\rho^2/(1-\rho^2)}{L}. \tag{10}$$

The hidden-phase sampler has marginal output $\overline{Q}$ satisfying

$$\chi^2(\overline{Q}\|P_n) \leq V_*, \qquad D_{\mathrm{KL}}(P_n\|\overline{Q}) \leq (2+8t)V_* + 4\sqrt{B_\rho t V_*/L}. \tag{11}$$

Every deterministic adaptive R-round policy satisfies

$$\inf_{\sigma\in\mathcal{D}_R} D_{\mathrm{KL}}(P_n\|Q_\sigma) \geq d_\rho t. \tag{12}$$

B.2 PROOFS FOR THE FIXED SYMMETRIC CHAIN

This appendix proves Theorem 5 and Corollaries 3 and 7. Throughout, expectations without another subscript are under P_n; $R, s \geq 1$ are integers, $L = 2^{R-1}$, $n = (s+1)L$, $u = \rho^L$, $v = u^2$, and $t = sv$. The upper-bound estimates use $u \leq 1/4$; the deterministic converse in Appendices B.2.5–B.2.8 holds without that restriction.

The proof first establishes the construction's marginal upper bound, then charges unavoidable same-round dependence to every deterministic policy. The asymptotic consequences and interpretation of phase mixing follow only after these two bounds are complete.

B.2.1 LEGALITY AND THE IMPLEMENTED DENSITY

For phase i, the $s+1$ first-round positions are $i + jL$, $0 \leq j \leq s$. The left tail has $i - 1 \leq L - 1$ sites and the right tail has $L - i \leq L - 1$ sites. Every interior gap has exactly $L - 1$ sites. An interval of at most $2^k - 1$ sites can be completed in k rounds by choosing a middle separator and recursing on the two remaining intervals. At any round the current intervals are conditionally independent given the revealed sites, by the Markov property. Sampling one separator per interval using singleton conditionals thus has no within-round factorization error. The recursion has the base case $k = 0$, when the interval is empty. Applying it with $k = R - 1$ proves legality and exact conditional completion, including both tails and $R = 1$.

The target marginal on these positions is a stationary binary chain with kernel $K^L(x, y) = (1 + uxy)/2$. Independent first-round draws replace that chain by its product uniform marginal. Since later completion is exact, their likelihood ratio is the full output likelihood ratio:

$$T_i(x) = \frac{Q_i(x)}{P_n(x)} = \prod_{j=0}^{s-1} \frac{1}{1 + u x_{i+jL} x_{i+(j+1)L}}.$$

Under the target, $Z_e = X_e X_{e+1}$ are independent mean-ρ signs: the map from X to $(X_1, Z_1, \ldots, Z_{n-1})$ is bijective and the density factorizes. This gives (6). In one phase the $Y_{i,j}$ are independent mean-u signs. In particular $\mathbb{E}T_i = 1$,

$$\mathbb{E}T_i^2 = (1 - v)^{-s}, \qquad \mathbb{E}[-\log T_i] = s\mathbb{E}\log(1 + uY) = s\phi(u).$$

The last identity follows from $\log(1 + uy) = \frac{1}{2}\log(1 - u^2) + y\,\mathrm{atanh}(u)$ for $y \in \{-1, +1\}$.

B.2.2 CROSS-PHASE TRANSFER AND ITS BOUNDARY TERMS

Fix $1 \leq h < L$ and phases $i, i + h$ within $\{1, \ldots, L\}$. Split the relevant edge windows into independent signs $A_0, B_0, A_1, B_1, \ldots, B_{s-1}, A_s$ with means $a = \rho^h$ and $b = \rho^{L-h}$ respectively. The two window sequences are $A_j B_j$ and $B_j A_{j+1}$. Since $ab = u$, integration over B_j gives the kernel

$$G(A, A') = \mathbb{E}_B \frac{1}{(1 + uAB)(1 + uBA')} = \frac{1 - ub(A + A') + u^2 AA'}{(1 - v)^2}.$$

Consider the operator $(\mathcal{G}f)(A) = \mathbb{E}_{A'} G(A, A') f(A')$ on linear functions $f(A) = c_0 + c_1 A$. Its coefficient matrix and boundary moment are

$$M_h = \frac{1}{(1 - v)^2} \begin{pmatrix} 1 - v & a(1 - b^2) \\ -ab^2(1 - a^2) & 0 \end{pmatrix}, \qquad c_s(h) = (1 \quad a)\, M_h^s \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{13}$$

The right boundary is the constant function 1 and the left boundary averages over a mean-a sign. Therefore

$$c_s(h) = \mathbb{E}T_i T_{i+h} = [1, a] M_h^s [1, 0]^\top, \quad c_0 = 1, \quad c_1 = \frac{1 - 2v + va^2}{(1 - v)^2}.$$

Its trace and determinant are

$$\vartheta = \frac{1}{1 - v}, \qquad \delta = \frac{v(1 - a^2)(1 - b^2)}{(1 - v)^4}.$$

Let the real eigenvalues be $\lambda_+ \geq \lambda_-$. The squared eigenvalue gap has numerator

$$(1 - v)^2 - 4v(1 - a^2)(1 - b^2) \geq (15/16)^2 - 1/4 = 161/256.$$

The denominator is $(1-v)^4 \le 1$, so the gap is at least $3/4$. Both eigenvalues are nonnegative and

$$0 \le \lambda_- \le \frac{2\delta}{\vartheta} \le \frac{2v}{(1-v)^3} \le \frac{1}{6}.$$

Write $k = a^2 + b^2$. Evaluating the characteristic polynomial at 1 gives

$$(1-\lambda_+)(1-\lambda_-) = \frac{v[-k+4v-3v^2+v^3]}{(1-v)^4} \le 0,$$

where $k \ge 2\sqrt{v} \ge 8v$. It follows that $\lambda_+ \ge 1$ and

$$\lambda_+ - 1 \le \frac{vk}{(1-v)^4(1-\lambda_-)} \le 2vk.$$

Also $1 \le c_1 \le 1 + 2va^2$. The recurrence solution is

$$c_s = \alpha\lambda_+^s + (1-\alpha)\lambda_-^s, \qquad \alpha = \frac{c_1 - \lambda_-}{\lambda_+ - \lambda_-} \ge 0.$$

If $\alpha \le 1$, $c_s \le \lambda_+^s$. Otherwise $\alpha - 1 = (c_1 - \lambda_+)/(\lambda_+ - \lambda_-) \le 3vk$, and the negative coefficient of λ_-^s can be dropped for an upper bound. For $s \ge 1$ this gives

$$c_s(h) \le (1+3vk)e^{2svk} \le e^{5tk}. \tag{14}$$

This calculation bounds the actual boundary moment, not just the spectral radius.

B.2.3 Variance of the averaged density

Using $\mathbb{E}T = 1$, the diagonal moment, and the $L-h$ ordered phase pairs at distance h gives

$$V := \mathrm{Var}_{P_n}(T) = \frac{(1-v)^{-s}-1}{L} + \frac{2}{L^2}\sum_{h=1}^{L-1}(L-h)(c_s(h)-1) \le V_*. \tag{15}$$

First, $-\log(1-v) \le 2v$ implies $(1-v)^{-s} \le e^{2t}$. Second, $k \le 2$ and $e^{5tk} - 1 \le 5te^{10t}k$. Finally,

$$\sum_{h=1}^{L-1}(\rho^{2h} + \rho^{2(L-h)}) \le \frac{2\rho^2}{1-\rho^2}.$$

Bounding $L - h \le L$ in the nonnegative upper estimates proves $V \le V_*$. Note that the individual $c_s(h)-1$ need not be assumed nonnegative to use their upper bounds. The exact total V is nonnegative by definition.

B.2.4 Auxiliary log cost and the KL lower tail

Across all phases, every starting position $k = 1, \ldots, sL$ appears once as a window start. Put $N_0 = sL$ and

$$S = \sum_{k=1}^{N_0} W_k, \qquad W_k = \prod_{e=k}^{k+L-1} Z_e, \qquad A = \frac{s}{2}\log(1-v) + \frac{\mathrm{atanh}(u)}{L}S.$$

For $1 \le h < L$, overlapping signs cancel in $W_k W_{k+h}$ and hence

$$\mathrm{Cov}(W_k, W_{k+h}) = \rho^{2h} - v;$$

for $h \ge L$ the covariance is zero. Thus the exact auxiliary variance is

$$\mathrm{Var}(S) = N_0(1-v) + 2\sum_{h=1}^{L-1}(N_0 - h)(\rho^{2h} - v) \le N_0 B_\rho, \tag{16}$$

$$\mathrm{Var}(A) = \frac{\mathrm{atanh}^2(u)}{L^2}\,\mathrm{Var}(S) \le \frac{4B_\rho t}{L}. \tag{17}$$

We used $\mathrm{atanh}(u) \le 2u$ for $u \le 1/4$. Moreover $\mathbb{E}A = s\phi(u) \le su\,\mathrm{atanh}(u) \le 2t$. AM–GM gives $-\log T \le A$ pointwise.

Let $f(z) = -\log z + z - 1$. Since $\mathbb{E}T = 1$, the desired KL is $\mathbb{E}f(T)$. For $z \geq 1/2$, Taylor's theorem with $f(1) = f'(1) = 0$ and $f''(z) = 1/z^2 \leq 4$ between z and 1 gives $f(z) \leq 2(z-1)^2$. For $\mathcal{H} = \{T < 1/2\}$, $f(T) \leq -\log T \leq A$ and $P_n(\mathcal{H}) \leq 4V$. Therefore

$$\mathbb{E}f(T) \leq 2V + \mathbb{E}[A\mathbf{1}_{\mathcal{H}}] \tag{18}$$

$$\leq 2V + (\mathbb{E}A)P_n(\mathcal{H}) + \sqrt{\mathrm{Var}(A)P_n(\mathcal{H})} \tag{19}$$

$$\leq (2 + 4\mathbb{E}A)V + 2\sqrt{\mathrm{Var}(A)V}. \tag{20}$$

Substituting the bounds for $\mathbb{E}A$, $\mathrm{Var}(A)$, and V proves (11). Centering A is essential for the displayed $O(1/L)$ conclusion at fixed t; a bound using only its uncentered second moment would lose that rate.

B.2.5 A SHARED HISTORY IDENTITY FOR DETERMINISTIC POLICIES

The argument in this subsection holds for any strictly positive target on a finite alphabet, not only the binary chain. Fix a deterministic policy σ. For any reachable history $h = (A, x_A)$, the set of full assignments that cause the policy to reach h is exactly the cylinder $\{X_A = x_A\}$. Indeed, the successive batches on the path are deterministic functions of the earlier revealed values, so every completion of those same values follows the same earlier decisions. No condition on unrevealed values or future scheduling choices is added by reaching the history.

For a full assignment x, let $h_r = (A_r, x_{A_r})$ and B_r be the histories and batches obtained by replaying the policy on x. Ratios of target cylinder probabilities telescope, whereas the sampler uses singleton products. Hence

$$\log \frac{P_n(x)}{Q_\sigma(x)} = \sum_r \log \frac{P_n(X_{B_r} = x_{B_r} \mid X_{A_r} = x_{A_r})}{\prod_{j \in B_r} P_n(X_j = x_j \mid X_{A_r} = x_{A_r})}. \tag{21}$$

All factors are positive. Taking the target expectation and conditioning at each history proves (4). These are histories weighted by the target law, not the generated law. Completed paths may be padded with zero cost. The identity requires batch selection before current-round samples and need not hold under candidate peeking or exposed extra sampling coins.

B.2.6 A SHORT WITNESS IN EVERY CELL

Lemma 6 (A short same-round witness)**.** *Every labeling of a path with 2^R vertices by labels in $\{1, \ldots, R\}$ contains two vertices with the same label, distance at most 2^{R-1}, and no smaller label strictly between them.*

Proof. For $R = 1$, the two adjacent vertices suffice. Suppose the result holds for $R - 1$. If two consecutive occurrences of label 1 have distance at most 2^{R-1}, they are the required pair. If two consecutive occurrences have larger distance, their interior contains a path of 2^{R-1} vertices without label 1, so induction applies after subtracting one from the remaining labels. If there is no occurrence of label 1, take any such subpath. If there is exactly one, one side contains at least 2^{R-1} vertices. These exhaust the possibilities. The inductive distance bound 2^{R-2} is at most the required 2^{R-1}. □

B.2.7 UNIFORM PAIR INFORMATION UNDER ARBITRARY BOUNDARY VALUES

For two unrevealed sites at distance d with no earlier revealed site between them, only the nearest revealed boundary on each side affects their conditional law. The binary transition powers give

$$p(x, y) = \frac{(1 + \alpha x)(1 + rxy)(1 + \beta y)}{4(1 + \alpha\beta r)}, \qquad r = \rho^d, \quad |\alpha|, |\beta| \leq \rho.$$

An absent boundary contributes field zero; a boundary at distance $m \geq 1$ and spin z contributes $z\rho^m$. Direct summation yields

$$\mathrm{Cov}(X, Y) = \frac{r(1 - \alpha^2)(1 - \beta^2)}{(1 + \alpha\beta r)^2}.$$

For two binary signs, the total variation between their joint distribution and the product of its marginals equals $|\mathrm{Cov}(X, Y)|/2$. Pinsker's inequality in nats therefore gives

$$I(X; Y) \geq \frac{(1 - \rho^2)^4}{2(1 + \rho^3)^4}\rho^{2L}, \qquad d \leq L. \tag{22}$$

The denominator bound uses $r \le \rho$ and $|\alpha\beta| \le \rho^2$. The information lower bound is uniform over every earlier revealed assignment.

B.2.8 A SHARED CHARGING RULE AND THE DETERMINISTIC LOWER BOUND

Partition the first $2L\lfloor n/(2L)\rfloor$ positions into fixed disjoint cells of length $2L = 2^R$. Initially mark every cell uncharged. At each history, before sampling its selected batch, inspect uncharged cells for two selected positions at distance at most L with no previously revealed position between them. Charge every qualifying cell and select one qualifying pair by a fixed tie-breaker.

The pairs selected at any node are disjoint. Grouping each such pair together and every remaining coordinate separately, entropy subadditivity gives

$$C(h) \ge \sum_{\text{pairs selected at } h} I(X_u; X_v \mid h).$$

Entropy subadditivity applies even when the selected pairs are dependent. Every complete path labels each cell by actual reveal rounds in $\{1, \dots, R\}$. Lemma 6 ensures that a qualifying pair is encountered unless its cell was charged earlier. Thus each cell is charged exactly once on every path; no current-round or future values are inspected.

This gives an alphabet-independent reduction. If every eligible pair has conditional mutual information at least η_L at every reachable history, then (4) gives

$$D_{\mathrm{KL}}(P_n \| Q_\sigma) \ge \eta_L \left\lfloor \frac{n}{2L} \right\rfloor. \tag{23}$$

For the symmetric chain, substitute $\eta_L = 2d_\rho \rho^{2L}$ from (22):

$$D_{\mathrm{KL}}(P_n \| Q_\sigma) \ge 2d_\rho \rho^{2L} \left\lfloor \frac{n}{2L} \right\rfloor = 2d_\rho v \left\lfloor \frac{s+1}{2} \right\rfloor \ge d_\rho s v.$$

This proves (12) for all deterministic adaptive policies with a hard cap, including policies with data-dependent stopping times below the cap.

B.2.9 ASYMPTOTIC CONSEQUENCES AND THE EXTRA ROUND

Corollary 7 (Vanishing versus divergent error)**.** *Fix $0 < \rho < 1$ and let $L \to \infty$ through powers of two, with $R = 1 + \log_2 L$ and $n = (s+1)L$. For any fixed $0 < \kappa < 1/10$, let $s = \lfloor \kappa \log L/\rho^{2L} \rfloor$. The randomized KL tends to zero and the infimum deterministic KL tends to infinity. Both classes use $R = \log_2 \log n + O_\rho(1)$ rounds. A deterministic $(R+1)$-round sampler already has vanishing KL on these sequences.*

For $t \le \tau$, $e^{2t} - 1 \le 2e^{2\tau} t$ and $te^{10t} \le e^{10\tau} t$, so $V_* \le C'_{\rho,\tau} t/L$. Equation (11) then gives $D_{\mathrm{KL}}(P_n \| \overline{Q}) \le C_{\rho,\tau} t/L$. When $s = \lfloor t_0/v \rfloor$, $t \to t_0 > 0$, proving Corollary 3.

For $s = \lfloor \kappa \log L/v \rfloor$, $t \sim \kappa \log L$. Since $\kappa < 1/10$,

$$V_* = O_\rho\big((\log L) L^{10\kappa - 1}\big), \quad (2+8t)V_* \to 0, \quad \sqrt{B_\rho t V_*/L} \to 0,$$

whereas $d_\rho t \to \infty$. Along either family $\log n = 2L|\log \rho| + O(\log L + |\log t|)$, so $R = \log_2 \log n + O_\rho(1)$.

For an $(R+1)$-round deterministic construction, use first-round positions of spacing $2L$ followed by R exact separator rounds. At most $O(s)$ transitions between consecutive first-round positions are factorized, each costing $\phi(\rho^{2L}) = O(\rho^{4L})$. Hence its total KL is $O(s\rho^{4L}) = O(tv) \to 0$ on both sequences.

B.2.10 INTERPRETATION: TAPE-CONDITIONED VERSUS MARGINAL COST

For completeness, let the target joint distribution on (I, X) be $\mu_i P_n(x)$ and the generated joint be $\mu_i Q_i(x)$. Their KL is $\sum_i \mu_i D_{\mathrm{KL}}(P_n \| Q_i)$. Applying the KL chain rule in the other order gives the marginal divergence plus $\mathbb{E}_{P_n} D_{\mathrm{KL}}(\mu \| \nu(\cdot \mid X))$, proving (5). Every output law is strictly positive; tape realizations with $\mu_i = 0$ contribute zero and may be omitted. In our construction the left side is $s\phi(u)$; it stays positive or diverges while the first term on the right vanishes. The posterior divergence quantifies the gain from hiding the phase.

B.2.11 INTERPRETATION: TRANSITION LOCATIONS, NOT THEIR COUNT

For the symmetric translated-grid construction, define the sign-change count $N(x) = \sum_{e=1}^{n-1} \mathbf{1}\{x_e \neq x_{e+1}\}$, and let $Q_w = \sum_{i=1}^{L} w_i Q_i$ for any fixed probability vector w. The distribution of N is identical under every component Q_i and Q_w. Moreover, at each count k, the conditional forward KL under Q_w is no greater than under any component.

To see this, use the edge signs $Z_e = X_e X_{e+1}$. Each component modifies their iid target law on s disjoint blocks of L edges, with the same factor $(1+uY)^{-1}$ on each block product Y. Each block is normalized: its target product probabilities $(1+uy)/2$ become $1/2$. The remaining $L-1$ tail edges retain their target law. Thus the component edge laws differ only by permutations of edge coordinates, which preserve N. Their common count law, denoted H_N, is also the count law of every mixture.

Write $P_k = P_n(\cdot \mid N = k)$ and $Q_{i,k} = Q_i(\cdot \mid N = k)$. The target P_k is uniform over initial signs and placements of k negative edges. The same permutations preserve P_k and map the component conditional laws into one another, so their conditional KLs have a common value d_k. Since the count law is common, conditioning does not reweight the phases: $Q_w(\cdot \mid N = k) = \sum_i w_i Q_{i,k}$. Convexity gives

$$D_{\mathrm{KL}}\left(P_k \middle\| \sum_i w_i Q_{i,k}\right) \leq d_k = D_{\mathrm{KL}}(P_k \| Q_{i,k}) \qquad \text{for every } i, k.$$

With P_N the target count law, the KL chain rule yields

$$D_{\mathrm{KL}}(P_n \| Q_i) - D_{\mathrm{KL}}(P_n \| Q_w) = \sum_{k=0}^{n-1} P_N(k) \left[d_k - D_{\mathrm{KL}}\left(P_k \middle\| \sum_j w_j Q_{j,k}\right)\right].$$

The count-marginal term $D_{\mathrm{KL}}(P_N \| H_N)$ cancels. The initial sign remains independent uniform under every law, so the entire gain concerns transition locations conditional on their count. This statement uses neither small u nor bounded t; it concerns these symmetric grid components, not arbitrary policies or learned conditionals.

B.3 PROOF OF THE RANDOMIZED MARGINAL FRONTIER

Theorem 8 (Randomized marginal lower bound)**.** *For the stationary zero-field chain $K(x,y) = (1+\rho xy)/2$, $0 < \rho < 1$, let $R, s \geq 1$ be integers, $L = 2^{R-1}$, $n = (s+1)L$, and $t = s\rho^{2L}$. Put*

$$\gamma_\rho = \frac{(1-\rho^2)^2}{(1+\rho^3)^2}, \qquad C_\rho = \frac{4(1+\rho)}{(1-\rho)^2}, \qquad c_\rho = \frac{\gamma_\rho^2}{4C_\rho^2} = \frac{(1-\rho)^8(1+\rho)^2}{64(1+\rho^3)^4}.$$

Every $\sigma \in \mathcal{R}_R$ satisfies, with $m = \lfloor n/(2L) \rfloor$,

$$D_{\mathrm{KL}}(P_n \| Q_\sigma) \geq \frac{2\gamma_\rho^2 \rho^{2L} m^2}{C_\rho^2 n} \geq c_\rho \frac{t}{L}. \tag{24}$$

The bound holds throughout $0 < \rho < 1$ and all positive integer R, s.

Write $P = P_n$, $v = \rho^{2L}$, and use m, γ_ρ, C_ρ, and c_ρ from the theorem. The proof uses a correlation statistic of the output: independent within-batch draws cause a predictable deficit, every reveal path incurs such deficits, and a concentration bound under the generated law turns them into a marginal KL lower bound. This does not average the deterministic KL bound.

Fix any policy in $\mathcal{R}_R$. Formally, it is driven by an exogenous scheduling tape I, independent of the oracle's private singleton sampling tape; all online scheduling randomness is included in I. Given (I, h), each next batch is fixed before any current-round value or private oracle coin is exposed; its values are sampled from the product of the exact target singleton conditionals. Every tape in the scheduling distribution's support and every positive-probability oracle-outcome path reveals all sites exactly once within R rounds. Completed paths are padded by empty batches. Let Q be its output law after averaging over the scheduling tape. Throughout, target conditional expectations mean conditioning only on the specified spins, not on a random future scheduling trace. Define

$$F_L(x) = \sum_{d=1}^{L} \rho^{L-d} \sum_{i=1}^{n-d} x_i x_{i+d}, \qquad \mu = \mathbb{E}_P F_L, \qquad M(A, x_A) = \mathbb{E}_P[F_L \mid X_A = x_A]. \tag{25}$$

B.3.1 NONNEGATIVE QUADRATIC CLOSURE AND CONDITIONAL DRIFT

For a fixed set C, every pair moment $g_{ij}(x_C) = \mathbb{E}_P[X_i X_j \mid X_C = x_C]$, $i < j$, has the form

$$g_{ij}(x_C) = a_0 + \sum_{k<\ell,\ k,\ell\in C} a_{k\ell} x_k x_\ell, \qquad a_{k\ell} \geq 0. \tag{26}$$

The coefficients depend on the positions, not the conditioning assignment; there are no terms of degree three or higher. To see this, note that $K^d(x, y) = (1 + \rho^d xy)/2$. For a single unrevealed site k between boundaries $l < k < r$, set $a = \rho^{k-l}$, $b = \rho^{r-k}$. Summing the bridge law and rationalizing its denominator gives

$$\mathbb{E}_P[X_k \mid x_l, x_r] = \frac{ax_l + bx_r}{1 + abx_l x_r} = \frac{a(1-b^2)}{1-a^2b^2} x_l + \frac{b(1-a^2)}{1-a^2b^2} x_r. \tag{27}$$

Both coefficients are nonnegative. With one boundary the mean is its spin times the corresponding power of ρ; with no boundary it is zero. A revealed spin is itself a linear expression with coefficient one.

Distinct unrevealed intervals are conditionally independent, so a pair in different intervals has moment equal to the product of these linear means. This produces only nonnegative quadratic terms, with squares absorbed into the constant. The same argument handles one or both sites revealed. If $l < i < j < r$ are within one interval, put $a = \rho^{i-l}$, $q = \rho^{j-i}$, $b = \rho^{r-j}$. Direct summation gives

$$g_{ij} = \frac{q + abx_l x_r}{1 + aqbx_l x_r} = \frac{q(1-a^2b^2)}{1-a^2q^2b^2} + \frac{ab(1-q^2)}{1-a^2q^2b^2} x_l x_r. \tag{28}$$

Its quadratic coefficient is nonnegative. With one or neither boundary, the same summation gives $g_{ij} = q$. This proves (26) for every fixed conditioning set, including arbitrary signed assignments.

We also need nonnegative conditional covariances. For two unrevealed sites $i < j$ with no revealed site between them, let $d = j - i$ and $q = \rho^d$. The nearest revealed boundary on each side induces a signed field α or β, of magnitude at most ρ; an absent boundary induces field zero. Their conditional law and covariance are

$$p(x, y) = \frac{(1 + \alpha x)(1 + qxy)(1 + \beta y)}{4(1 + \alpha\beta q)}, \tag{29}$$

$$\mathrm{Cov}_P(X_i, X_j \mid x_A) = \frac{q(1-\alpha^2)(1-\beta^2)}{(1+\alpha\beta q)^2} \geq \gamma_\rho q. \tag{30}$$

These are the conditional pair formulas from Appendix B.2.7; $|\alpha|, |\beta| \leq \rho$ and $q \leq \rho$ give the displayed inequality. Sites separated by a revealed site are conditionally independent, hence have zero covariance. Thus all new–new conditional covariances are nonnegative, for every history assignment.

Fix a history (A, x_A) and its chosen batch B, and put $q_B = \prod_{j\in B} P(X_j \mid X_A = x_A)$. Define its predictable drift by

$$D(A, B, x_A) = M(A, x_A) - \mathbb{E}_{q_B} M(A \cup B, x_A, X_B). \tag{31}$$

Apply (26) to every summand of F_L, conditioning on $A \cup B$. The true conditional law of X_B and q_B agree on constants and terms involving at most one new spin. Their expectation difference for a new–new term is its nonnegative coefficient times the conditional covariance. The target tower property identifies the first expectation with $M(A, x_A)$. More explicitly, write the expansion obtained by summing (26) with the weights in F_L as

$$M(C, x_C) = \beta_0(C) + \sum_{k<\ell,\ k,\ell\in C} \beta_{k\ell}(C) x_k x_\ell, \qquad \beta_{k\ell}(C) \geq 0, \quad C = A \cup B.$$

Then the exact drift is

$$D(A, B, x_A) = \sum_{k<\ell,\ k,\ell\in B} \beta_{k\ell}(A \cup B)\, \mathrm{Cov}_P(X_k, X_\ell \mid X_A = x_A) \geq 0. \tag{32}$$

For $i, j \in B$ with $j - i \leq L$, their own summand in F_L supplies $\beta_{ij}(A \cup B) \geq \rho^{L-(j-i)}$. Moreover, for any collection $\mathcal{W}$ of distinct pairs in B with $j - i \leq L$ and no point of A between them,

$$D(A, B, x_A) \geq \sum_{(i,j)\in\mathcal{W}} \rho^{L-(j-i)}\, \mathrm{Cov}_P(X_i, X_j \mid x_A) \geq |\mathcal{W}| \gamma_\rho \rho^L. \tag{33}$$

For each selected pair, its own summand in F_L becomes exactly $X_i X_j$ after conditioning on $A \cup B$; every other summand still has nonnegative drift. Distinct summands are counted once, so even overlapping selected pairs require no independence assumption.

B.3.2 Online charging on every adaptive path

Lemma 6 is a pathwise combinatorial statement, so it applies after fixing any scheduling tape and realized output. Reusing it here does not invoke a deterministic-KL lower bound.

Apply the online rule from Appendix B.2.8 to the m cells of length $2L$. Charges depend only on the old history and selected positions; every tape/output path charges each cell once. In the present argument, the cost of a charge is the conditional drift in (33), not pair mutual information. Thus

$$\sum_{r=0}^{R-1} D_r \geq \delta, \qquad \delta = m\gamma_\rho \rho^L, \qquad \text{on every tape/output path.} \tag{34}$$

The pathwise labeling verifies coverage; it is never used as conditioning information when evaluating a covariance or choosing a charge.

B.3.3 Uniform oscillation under a single revealed-spin change

Flipping a single coordinate of F_L changes at most two pairs per distance, each by at most twice its weight. Its coordinate oscillation is therefore at most

$$b_\rho = 4 \sum_{d=1}^{L} \rho^{L-d} \leq \frac{4}{1-\rho}.$$

Fix a conditioning set C, a site $j \in C$, and arbitrary values on $C \setminus \{j\}$. Write $x_C^{j,+}$ and $x_C^{j,-}$ for the assignments with those same values and spin $+1$ or -1 at j, respectively. Couple the two target conditional laws with $x_j = -1$ and $x_j = +1$. All intervals not adjacent to j have identical laws and are coupled identically. We now construct a joint monotone coupling of each affected interval, not just couplings of its one-site marginals.

For the interval to the right, sample away from j towards its unchanged right boundary r with fixed sign z. At the next site $k < r$, given previous spin s, the conditional mean follows (27) with $a = \rho$ and $b = \rho^{r-k}$:

$$H_k(s, z) = \frac{\rho(1-b^2)}{1-\rho^2 b^2} s + \frac{b(1-\rho^2)}{1-\rho^2 b^2} z.$$

The coefficient of s is nonnegative for *either* $z = -1$ or $z = +1$. At each step use a common uniform variable and output $+1$ when it is at most $(1 + H_k(s, z))/2$. Starting from ordered endpoints, the two paths stay ordered; each has the correct sequential bridge law. If the other boundary is absent, the transition mean is simply ρs. On the left interval use the same construction in reverse spatial order: the symmetric kernel and uniform stationary law make the reversed chain identical. This constructs the coupling for arbitrary signed unchanged boundaries on both sides. Other revealed sites remain identical.

At a site k distance d from j, with the next unchanged boundary a further $e \geq 1$ sites away, order implies that the disagreement probability is half the difference of the two marginal means. By (27) it equals

$$\frac{\rho^d(1-\rho^{2e})}{1-\rho^{2(d+e)}} \leq \rho^d.$$

Without the other boundary it is exactly ρ^d. At j it is one, and at other revealed sites it is zero. Hence the expected Hamming distance is at most $1 + 2\sum_{d \geq 1} \rho^d = (1+\rho)/(1-\rho)$. Telescoping single-site changes of F_L along the coupled configurations proves

$$|M(C, x_C^{j,+}) - M(C, x_C^{j,-})| \leq b_\rho \frac{1+\rho}{1-\rho} \leq C_\rho. \tag{35}$$

This bound is uniform in the set C, its values, n, and L.

B.3.4 THE COMPENSATED SUPERMARTINGALE FOR RANDOM BATCH SIZES

Include the entire independent scheduling tape I in $\mathcal{F}_0$; fresh external coins can equivalently be sampled in advance. Let $\mathcal{F}_r$ contain I and the complete history after r rounds. Write A_r for the revealed set after r rounds and B_r for the batch selected for round $r+1$, so $A_{r+1} = A_r \cup B_r$ and $A_0 = \varnothing$. Then B_r is $\mathcal{F}_r$-measurable, and conditional on $\mathcal{F}_r$ its spins have law q_{B_r}. Denote the joint tape/output law by $\widetilde{Q}$ and use it for the conditional expectations below. Set $M_r = M(A_r, X_{A_r})$ and

$$\eta_r = M_{r+1} - \mathbb{E}_{\widetilde{Q}}[M_{r+1} \mid \mathcal{F}_r], \qquad D_r = M_r - \mathbb{E}_{\widetilde{Q}}[M_{r+1} \mid \mathcal{F}_r].$$

These are exactly the drifts in (31), and $M_{r+1} - M_r = \eta_r - D_r$. Pad paths finishing early with empty batches and zero increments after completion.

Given $\mathcal{F}_r$, expose the independent values at positions in B_r one by one solely to analyze the fixed function $M(A_r \cup B_r, x_{A_r}, X_{B_r})$; the policy does not observe these partial exposures. Each Doob difference has conditional mean zero and conditional range length at most C_ρ by (35): averaging over the remaining independent spins preserves this oscillation bound. For a centered variable Y with range length at most C_ρ, Hoeffding's bound is $\log \mathbb{E}e^{\lambda Y} \le \lambda^2 C_\rho^2/8$. For completeness, the log-mgf has value and derivative zero at the origin; its second derivative is the variance under exponential tilting, at most $C_\rho^2/4$, since the tilted support stays in the same interval. Integrating twice gives the bound. Iterating over the batch yields

$$\mathbb{E}_{\widetilde{Q}}[e^{\lambda \eta_r} \mid \mathcal{F}_r] \le \exp\{a_\lambda |B_r|\}, \qquad a_\lambda = \lambda^2 C_\rho^2/8. \tag{36}$$

To account for adaptive batch sizes, let $N_k = \sum_{r<k} |B_r|$ and define the nonnegative compensated process

$$Z_k = \exp\left\{\lambda \sum_{r<k} \eta_r - a_\lambda N_k\right\}, \qquad Z_0 = 1.$$

Since $|B_k|$ is known at $\mathcal{F}_k$, (36) gives $\mathbb{E}_{\widetilde{Q}}[Z_{k+1} \mid \mathcal{F}_k] \le Z_k$. Iterate for the fixed padded horizon R. The reveal-once rule implies $N_R = n$ on every path, so

$$\mathbb{E}_{\widetilde{Q}} \exp\{\lambda \textstyle\sum_{r<R} \eta_r\} \le e^{a_\lambda n}.$$

Now $M_0 = \mu$, $M_R = F_L$ and $F_L - \mu = \sum_{r<R}(\eta_r - D_r)$. Combining with (34), for $\lambda \ge 0$,

$$\mathbb{E}_Q e^{\lambda(F_L - \mu)} \le \exp\{-\lambda\delta + \lambda^2 C_\rho^2 n/8\}. \tag{37}$$

Here the left side is already marginal: F_L is a function of the output alone. Equivalently, the same proof conditional on I gives a uniform bound for its conditional output law Q_I; integrating $\mathbb{E}_{Q_I} e^{\lambda(F_L-\mu)}$ against *any* tape prior gives (37).

B.3.5 KL LOWER BOUND AND MATCHING UPPER BOUND

For positive laws on the finite output space and any real function f, Jensen's inequality gives

$$\log \mathbb{E}_Q e^f = \log \mathbb{E}_P \exp\{f + \log(Q/P)\} \ge \mathbb{E}_P f - D_{\mathrm{KL}}(P\|Q).$$

The target is strictly positive, and so are the generated laws: for a fixed legal tape and any proposed full assignment, all successive singleton draw probabilities along its adaptive path are positive. Taking $f = \lambda(F_L - \mu)$ and using (37) yields

$$D_{\mathrm{KL}}(P\|Q) \ge \lambda\delta - \lambda^2 C_\rho^2 n/8.$$

The optimizer $\lambda_* = 4\delta/(C_\rho^2 n)$ is nonnegative. Thus

$$\begin{aligned} D_{\mathrm{KL}}(P\|Q) &\ge \frac{2\delta^2}{C_\rho^2 n} = \frac{2\gamma_\rho^2 v m^2}{C_\rho^2 n} \\ &\ge \frac{\gamma_\rho^2 s^2 v}{2C_\rho^2 (s+1) L} \ge \frac{\gamma_\rho^2}{4C_\rho^2}\frac{t}{L}, \end{aligned}$$

where $m = \lfloor (s+1)/2 \rfloor \geq s/2$ for integer s and $s/(s+1) \geq 1/2$. Finally $c_\rho = (1-\rho)^8(1+\rho)^2/[64(1+\rho^3)^4]$, as stated. This proves Theorem 8 for all positive integer R, s.

To finish Theorem 1, combine this converse with (11), valid when $\rho^L \leq 1/4$. For $0 < t \leq \tau < \infty$, the quantity V_* in (10) satisfies

$$V_* \leq A_{\rho,\tau}\frac{t}{L}, \qquad A_{\rho,\tau} = 2e^{2\tau} + \frac{20e^{10\tau}\rho^2}{1-\rho^2}.$$

Thus one may take $C_{\rho,\tau} = (2+8\tau)A_{\rho,\tau} + 4\sqrt{B_\rho A_{\rho,\tau}}$. The hidden-phase sampler belongs to $\mathcal{R}_R$, while the lower bound holds for every member. Taking the infimum proves the stated order comparison on all feasible integer parameters in the stated regime.

B.4 The marginal converse for biased binary chains

The full frontier extends to unequal stationary state frequencies. Its upper bound and deterministic order use Theorem 10, proved in Appendix D.2; the proof below establishes the matching randomized marginal converse.

Corollary 9 (Sharp frontier with unequal state frequencies)**.** *Let $\pi = (1-p, p)$, $0 < p < 1$, and $K = \rho I + (1-\rho)\mathbf{1}\pi$ with $0 < \rho < 1$. Let P_n be its stationary path law. Under the parameter conditions of Theorem 10,*

$$\inf_{\sigma \in \mathcal{D}_R} D_{\mathrm{KL}}(P_n \| Q_\sigma) = \Theta_K(t), \qquad \inf_{\sigma \in \mathcal{R}_R} D_{\mathrm{KL}}(P_n \| Q_\sigma) = \Theta_{K,\tau}(t/L).$$

Let $\pi = (q, p)$ with $q = 1-p$ and $0 < p \leq 1/2$, and let $K = \rho I + (1-\rho)\mathbf{1}\pi$ for $0 < \rho < 1$. Let P_n be the stationary path law with initial distribution π and kernel K. Let R, s be positive integers, using the exact-singleton, reveal-once interface of Section 3. All target conditionals below condition only on the specified coordinate values. The state 1 is thus the less frequent state. Every strictly positive binary kernel with positive nonconstant eigenvalue has this form after relabelling. The proof follows the symmetric marginal converse, replacing spin symmetry by nonnegative bridge polynomials and association; the concentration argument still applies to every policy in $\mathcal{R}_R$.

Put $L = 2^{R-1}$, $n = (s+1)L$, and $t = s\rho^{2L}$ as before. In this appendix $X_i \in \{0, 1\}$ denotes the raw rare-state indicator. Define

$$F_L(x) = \sum_{d=1}^{L} \rho^{L-d} \sum_{i=1}^{n-d} x_i x_{i+d}, \qquad M(A, x_A) = \mathbb{E}_P[F_L \mid X_A = x_A], \qquad \mu = \mathbb{E}_P F_L.$$

Unlike a centered indicator, the raw indicator preserves nonnegative coefficients in the bridge moments. The conditional potential is now multilinear rather than quadratic; association supplies the required drift.

B.4.1 Nonnegative bridge polynomials

We call a revealed coordinate together with its fixed conditioning value a *pin*; an unpinned interval consists of unrevealed coordinates. Write $K_v(x, y) = \pi(y) + v(\mathbf{1}_{x=y} - \pi(y))$, so $K^d = K_{\rho^d}$. For two sites $l < i \leq j < r$ between their nearest pins, set $a = \rho^{i-l}$, $c = \rho^{j-i}$, $b = \rho^{r-j}$, $z = acb$, and $h = p + qc$. The case $i = j$, hence $c = 1$, gives a single-site mean. The pair moment is

$$g(x, y) = \mathbb{E}[X_i X_j \mid X_l = x, X_r = y] = \frac{K_a(x, 1)K_c(1, 1)K_b(1, y)}{K_z(x, y)} = A_0 + A_1 x + A_2 y + A_{12} xy,$$

where direct subtraction of the four boundary values gives

$$A_0 = \frac{pq(1-a)(1-b)h}{q+pz},$$
$$A_1 = \frac{a(1-b)h(q+pbc)}{(1-z)(q+pz)}, \qquad A_2 = \frac{b(1-a)h(q+pac)}{(1-z)(q+pz)},$$
$$A_{12} = \frac{abh\big[(1-c)(q+pz) + (1-2p)c(1-a)(1-b)\big]}{(1-z)(q+pz)(p+qz)}.$$

All four coefficients are nonnegative. With only a left pin, the moment is $[p(1-a)+ax]h$; the right-pin expression follows by reversibility. Without pins it is ph. Conditional independence between distinct unpinned intervals multiplies their single-site mean polynomials. A pinned site contributes its own indicator. Absorbing powers by $x^k = x$ therefore shows that, for any fixed conditioning set C, every conditional pair moment is a multilinear polynomial in x_C with nonnegative coefficients and degree at most four. Fixing any subset of its arguments to 0 or 1 preserves this property.

B.4.2 Association and the predictable deficit

The target conditioned on arbitrary pins is associated: increasing functions of its coordinates have nonnegative covariance. Here is a direct proof covering all pin assignments. Inside an unpinned interval, the next transition, given the preceding state x, has rows proportional to $K(x,y)r(y)$, where the positive function r accounts for the fixed future boundary. The determinant is positive, so its probability of state 1 is nondecreasing in x. Sequential sampling with shared thresholds thus expresses the interval as increasing functions of independent uniforms; its first site can be sampled from its appropriate initial law. Independent coordinates are associated by induction and the identity

$$\mathrm{Cov}(f(U), g(U)) = \tfrac{1}{2}\mathbb{E}[(f(U)-f(U'))(g(U)-g(U'))] \geq 0$$

for increasing one-variable functions and an independent copy U'. The induction uses total covariance, since averaging an increasing function over some coordinates preserves monotonicity in the others. Increasing maps preserve association, and independent intervals can use disjoint uniforms. Pins are constants. This proves the assertion.

Consequently, for any subset S of new sites and any pinned history,

$$\mathbb{E}_P\left[\prod_{j\in S} X_j \mid x_A\right] \geq \prod_{j\in S} \mathbb{E}_P[X_j \mid x_A].$$

This follows by induction, since products of nonnegative increasing indicators are increasing. For a selected batch B, let $q_B = \prod_{j\in B} P(X_j \mid x_A)$. Apply this inequality term by term to the nonnegative polynomial $M(A\cup B, x_A, X_B)$. The target tower property gives

$$D(A,B,x_A) := M(A,x_A) - \mathbb{E}_{q_B} M(A\cup B, x_A, X_B) \geq 0.$$

For any pair $i,j \in B$ at distance $d \leq L$, its direct summand in F_L is already fixed after the update. Its contribution to D is exactly $\rho^{L-d}\,\mathrm{Cov}_P(X_i, X_j \mid x_A)$. Every other summand has a nonnegative expectation deficit by the same polynomial argument. Thus any collection of distinct such pairs can be charged simultaneously.

B.4.3 Covariance and oscillation bounds uniform in the pins

Let $m_0 = 1-\rho$, $M_{\max} = 1/p$, and define fixed-kernel constants

$$\gamma = \frac{pq\, m_0^4}{M_{\max}^4}, \qquad H_{\mathrm{bin}} = 1 + 2(M_{\max}/m_0)^2 \frac{\rho}{1-\rho}, \qquad C_0 = \frac{2H_{\mathrm{bin}}}{1-\rho}.$$

For unrevealed sites $i<j$ with no pin between them, their conditional pair law is a product tilt of $H_{xy} = \pi(x)K_{\rho^{j-i}}(x,y)$:

$$P(X_i = x, X_j = y \mid x_A) = Z^{-1} l(x) H_{xy} r(y).$$

A left boundary contributes $l(x) = K_a(w,x)/\pi(x)$ and a right boundary contributes $r(y) = K_b(y,w')/\pi(w')$, with $a,b \leq \rho$; an absent boundary contributes 1. Every such factor lies in $[m_0, M_{\max}]$. Since $\det H = pq\rho^{j-i}$, $Z \leq M_{\max}^2$, and a Bernoulli pair's covariance equals the determinant of its probability table,

$$\mathrm{Cov}_P(X_i, X_j \mid x_A) \geq \gamma\rho^{j-i}.$$

Pairs separated by a pin are independent.

The online cell argument in Appendix B.3 is purely geometric: each disjoint cell of $2L$ consecutive sites forces a same-round pair at distance at most L with no earlier pin between them. Applying the preceding deficit bound to these distinct pairs yields, on every path,

$$\sum_{r<R} D_r \geq \delta := \lfloor n/(2L)\rfloor \gamma\rho^L.$$

To control fluctuations, flip one pin j while retaining all other pins. Couple each adjacent interval in spatial order with shared uniforms as above, using reversibility for the left interval. The paths stay ordered. At a site distance d from j, write r for the normalized factor of the other boundary, or $r \equiv 1$ if absent. The difference between its two conditional probabilities of state 1 is

$$\frac{\rho^d r(0) r(1)}{(K_{\rho^d} r)(0)(K_{\rho^d} r)(1)} \le (M_{\max}/m_0)^2 \rho^d.$$

The equality follows by subtracting the two tilted transition probabilities; the determinant of K_{ρ^d} is ρ^d. Order makes this difference the disagreement probability. All nonadjacent intervals can be coupled identically. The expected Hamming distance of the complete configurations is at most H_{bin}. One coordinate changes F_L by at most $2/(1-\rho)$, so telescoping coordinate changes in this coupling gives

$$|M(C, x_C^{j,1}) - M(C, x_C^{j,0})| \le C_0$$

uniformly in C, its pin values, and the path length.

B.4.4 Concentration under the generated marginal law

Let $\widetilde{Q}$ denote the joint law of the scheduling tape and generated output. Include the tape, independent of the private singleton sampling coins, in the initial filtration and pad completed paths with empty batches. With the notation of Appendix B.3, $M_{r+1} - M_r = \eta_r - D_r$. Conditional on the current history, each chosen batch has independent coordinates, and each coordinate has oscillation at most C_0 in the next potential. The same blockwise Hoeffding argument therefore gives

$$\mathbb{E}_{\widetilde{Q}}[e^{\lambda \eta_r} \mid \mathcal{F}_r] \le e^{\lambda^2 C_0^2 |B_r|/8}.$$

The compensated process using the cumulative revealed count is a supermartingale. Reveal-once completion fixes that count at n, even for adaptive batch sizes. Thus, for every $\lambda \ge 0$,

$$\mathbb{E}_Q e^{\lambda(F_L - \mu)} \le e^{-\lambda\delta + \lambda^2 C_0^2 n/8}.$$

This is an inequality for the returned law: F_L does not depend on the hidden tape. The KL variational inequality and optimization in λ give

$$D_{\mathrm{KL}}(P_n \| Q) \ge \frac{2\delta^2}{C_0^2 n} \ge \frac{\gamma^2}{4C_0^2} \frac{t}{L},$$

using $\lfloor (s+1)/2 \rfloor \ge s/2$ and $s/(s+1) \ge 1/2$. This lower bound holds for all $R, s \ge 1$; the bounded-t and small-ρ^L conditions are needed only for the matching upper bound.

C The first-set entropy frontier

We prove Theorem 2 by refining the marginal converse of Appendix B.3. First weight a fixed collection of witnesses by first-set probabilities; their collision probability then gives the entropy-sensitive lower bound. Combine it with the universal t/L floor and match both terms by mixing a controlled number of phases. Write $P = P_n$, $u = \rho^L$, $t = su^2$, and

$$m = \left\lfloor \frac{n}{2L} \right\rfloor \ge \frac{s}{2}, \qquad H = \frac{1+\rho}{1-\rho}, \qquad \gamma = \frac{(1-\rho^2)^2}{(1+\rho^3)^2}.$$

Fix a legal policy. Its distinct possible first sets have positive probabilities $w_1, \dots, w_N$, with $\sum_i w_i = 1$. Set $c_2 = \sum_i w_i^2$ and $w_* = \max_i w_i$. There is no restriction on N beyond finiteness, or on subsequent legal scheduling randomness.

C.1 A fixed cover of adaptive witnesses

Partition the first $2Lm$ sites into m cells of length $2L$. For $R \ge 2$, each cell has either a first-round pair at distance at most L, or a later same-round pair at distance at most $L/2$ with no earlier revealed site between its endpoints. Indeed, if the first alternative fails, the cell contains at most two first-round sites. Two such sites have a gap exceeding L; with one site, one side has at least L sites. In every

case there is an L-site subinterval with no first-round site. Apply Lemma 6 to its reveal labels in $\{2, \dots, R\}$, subtracting one from each label. The resulting pair has distance at most $2^{R-2} = L/2$ and no smaller label between its endpoints. This argument holds on every adaptive path.

For each possible first set i and each cell C, choose a pair $e_i(C)$ of distance at most L in that first set if one exists. Let E_{short} contain all within-cell pairs of distance at most $L/2$, and let E be its union with all the designated pairs, removing duplicates. When $R = 1$, put $E_{\mathrm{short}} = \varnothing$; all positions must then be sampled in the first round. For an edge $e = (j, k)$ of distance $d = k - j$, put $a_e = \rho^{L-d}$ and

$$v_e = \max\big(\{w_i : e = e_i(C) \text{ for some } C\} \cup \{w_* : e \in E_{\mathrm{short}}\}\big).$$

The defining set is nonempty for every $e \in E$. On every path with first set i, each cell supplies an eligible pair with $v_e \geq w_i$. For $R \geq 2$, the shorter pairs in one cell have total weight at most

$$\sum_{d=1}^{L/2} (2L - d)\rho^{L-d} \leq \frac{2L\rho^{L/2}}{1-\rho} \leq \frac{4\rho}{(1-\rho)^2} = H^2 - 1.$$

The second inequality uses $k\rho^k \leq \sum_{j=1}^{k} \rho^j \leq \rho/(1-\rho)$ with $k = L/2$. Since a squared maximum is at most the sum of squares and $w_*^2 \leq c_2$, the fixed cover satisfies

$$W_2 := \sum_{e \in E} a_e v_e^2 \leq mc_2 + m(H^2 - 1)w_*^2 \leq mH^2 c_2. \tag{38}$$

C.2 From marginal concentration to collision probability and entropy

Set $F(x) = \sum_{(j,k)\in E} a_{(j,k)} v_{(j,k)} x_j x_k$ and $\mu = \mathbb{E}_P F$. Include the entire independent scheduling tape in the filtration $\mathcal{F}_r$ after round r, and let $U_r = \mathbb{E}_P[F \mid X_{A_r}]$, using the specified revealed spins only. Under the joint scheduling/output law $\widetilde{Q}$, define

$$D_r = U_r - \mathbb{E}_{\widetilde{Q}}[U_{r+1} \mid \mathcal{F}_r], \qquad \eta_r = U_{r+1} - \mathbb{E}_{\widetilde{Q}}[U_{r+1} \mid \mathcal{F}_r].$$

The nonnegative quadratic closure (26) and drift identity (32) apply term by term. All drifts are nonnegative. By (30), an eligible pair of distance d contributes at least $\rho^{L-d} v_e \gamma \rho^d = \gamma u v_e$. Charging each cell at its first eligible pair from E whose coefficient v_e is at least w_i gives $\sum_{r<R} D_r \geq d_0 w_i$ on every path with first set i, where $d_0 = m\gamma u$.

A coordinate oscillation bound is $b_j = 2\sum_{e \ni j} a_e v_e$. At most two pairs of each distance meet a site, so weighted Cauchy–Schwarz gives

$$\sum_j b_j^2 \leq 4 \sum_j \Big(\sum_{e \ni j} a_e\Big)\Big(\sum_{e \ni j} a_e v_e^2\Big) \leq \frac{16 W_2}{1-\rho}.$$

The joint bridge coupling used for (35) bounds the oscillation of any target conditional mean in a revealed coordinate j by $h_j = \sum_k \rho^{|j-k|} b_k$. Zero extension and the convolution bound imply

$$S := \sum_j h_j^2 \leq H^2 \sum_j b_j^2 \leq \frac{16 H^4 m c_2}{1-\rho}. \tag{39}$$

Given the full tape and history, values in the selected batch are sampled independently from the exact singleton conditionals. The bounded-difference argument of (36) gives $\mathbb{E}[e^{\lambda \eta_r} \mid \mathcal{F}_r] \leq \exp\{\lambda^2 \sum_{j \in B_r} h_j^2/8\}$. Compensating by these predictable site costs yields deterministic terminal cost S, since each site is revealed once. Using $F - \mu = \sum_r (\eta_r - D_r)$, first condition on the full tape and then integrate within each first-set class. The actual output marginal obeys

$$\mathbb{E}_Q e^{\lambda(F-\mu)} \leq e^{\lambda^2 S/8} \sum_i w_i e^{-\lambda d_0 w_i}, \qquad \lambda \geq 0.$$

For $I \sim w$, $\mathbb{E} w_I = c_2$ and its range has length at most w_*. Hoeffding's lemma therefore bounds the right side by $\exp\{-\lambda d_0 c_2 + \lambda^2 (S + d_0^2 c_2)/8\}$. The KL variational inequality and optimization yield

$$D_{\mathrm{KL}}(P\|Q) \geq \frac{2 d_0^2 c_2^2}{S + d_0^2 c_2} \geq \frac{\kappa_\rho t}{1 + \kappa_\rho t} c_2, \qquad \kappa_\rho = \frac{(1-\rho)^9}{16(1+\rho^3)^4}. \tag{40}$$

Here $s/2 \leq m \leq s$. This lower bound does not require small u or bounded t. Concavity of $\log$ gives $-H_0 = \sum_i w_i \log w_i \leq \log c_2$. Thus $H_0 \leq h$ and $t \leq \tau$ imply $D_{\mathrm{KL}}(P\|Q) \geq [\kappa_\rho/(1 + \kappa_\rho \tau)] t e^{-h}$. Combining with (24) proves the lower half of (3).

C.3 Matching upper bound from a limited number of phases

Put $q = \min\{L, \lfloor e^h \rfloor\}$ and choose any q distinct phases $\mathcal{I} \subseteq \{1, \dots, L\}$ uniformly. The construction in Appendix B.2.1 is legal and has exactly q possible first sets. Its first-set entropy is $\log q \le h$, and $q^{-1} \le 2\max\{e^{-h}, L^{-1}\}$. Its output density is $T = q^{-1}\sum_{i\in\mathcal{I}} T_i$, where T_i is given by (6). Write $v = u^2$, $B_\rho = (1+\rho^2)/(1-\rho^2)$, and $k_h = \rho^{2h} + \rho^{2(L-h)}$. The finite phase index set obeys

$$\sum_{j\neq i} k_{|i-j|} = \sum_{h=1}^{L-1} k_h \le B_\rho - 1, \qquad 1 \le i \le L,$$

where the equality uses $k_h = k_{L-h}$. Restricting a row to $\mathcal{I}$ only decreases this sum. The diagonal moment in Appendix B.2.1 and the cross-phase estimate (14) imply, for $u \le 1/4$ and $t \le \tau$,

$$V := \mathbb{E}_P(T-1)^2 \le \frac{e^{2t}-1}{q} + \frac{5te^{10t}}{q^2}\sum_{\substack{i,j\in\mathcal{I}\\ i\neq j}} k_{|i-j|} \le K_\tau B_\rho \frac{t}{q}, \qquad K_\tau = 5e^{10\tau}.$$

To control the forward-KL lower tail, set $G = q^{-1}\sum_{i\in\mathcal{I}}\sum_{j=0}^{s-1} Y_{i,j}$ and

$$A = -\frac{1}{q}\sum_{i\in\mathcal{I}} \log T_i = \frac{s}{2}\log(1-v) + \operatorname{atanh}(u)G.$$

For phases at distance h, their window sums have covariance $s(\rho^{2h} - v) + (s-1)(\rho^{2(L-h)} - v)$: the overlapping windows have start distances h and $L-h$, respectively. Within a phase the windows are independent. The same row bound therefore gives $\mathrm{Var}_P(G) \le sB_\rho/q$. As in Appendix B.2.4, $\mathbb{E}_P A = s\phi(u) \le 2t$ and $\mathrm{Var}_P(A) \le 4B_\rho t/q$. AM–GM gives $-\log T \le A$, so the centered lower-tail argument (20) applies and yields

$$\begin{aligned} D_{\mathrm{KL}}(P\|Q) &\le (2+8t)V + 4\sqrt{B_\rho t V/q} \\ &\le B_\rho\big[(2+8\tau)K_\tau + 4\sqrt{K_\tau}\big]\frac{t}{q}. \end{aligned}$$

Together with the entropy lower bound, this proves (3). For the support statement, at most M first sets implies $H_0 \le \log M$, so the entropy lower gives order $t/\min\{M, L\}$. Choosing $\min\{M, L\}$ phases uniformly attains the same order.

D Distributional separation for finite-state chains

D.1 Finite-state bounds and consequences

For every dependent strictly positive reversible chain, the hidden-phase construction gives the stated finite-state output-law upper bound and a TV separation from the full deterministic class. For binary chains with positive correlation, its error order is optimal regardless of the stationary state frequencies.

Theorem 10 (Positive reversible chains)**.** *Fix a strictly positive reversible kernel K with stationary law π on a finite alphabet, and assume $K(x,y) \not\equiv \pi(y)$. Let P_n be its stationary path law, $\rho = \max_{j\ge 2}|\lambda_j(K)| \in (0,1)$, and $B = 1/\min_x \pi(x)$. Use the interface of Section 3. For integers $R, s \ge 1$, set $L = 2^{R-1}$, $n = (s+1)L$, and $t = s\rho^{2L}$. Then*

$$c_K t \le \inf_{\sigma\in\mathcal{D}_R} D_{\mathrm{KL}}(P_n\|Q_\sigma) \le Bt, \tag{41}$$

where $c_K > 0$ depends only on K. For fixed $0 < \tau < \infty$, if $0 < t \le \tau$ and $B\rho^L \le 1/2$, the uniform hidden-phase sampler with independent π-distributed first-round values satisfies

$$\chi^2(\overline{Q}\|P_n) \le C_{K,\tau}\, t/L, \qquad D_{\mathrm{KL}}(P_n\|\overline{Q}) \le C'_{K,\tau}\, t/L, \tag{42}$$

with finite constants $C_{K,\tau}, C'_{K,\tau}$ independent of R, s.

For $s = \lfloor t_0/\rho^{2L} \rfloor$ and fixed $t_0 > 0$, the deterministic optimum stays positive while the hidden-phase error vanishes as $L \to \infty$.

Corollary 9 combines this upper bound with the binary marginal converse to cover unequal stationary state frequencies. The general reversible upper bound also allows negative and repeated modes (Appendix D.2). A matching randomized converse extends to independent products $K = K_Y \otimes K_Z$, where K_Y is as in Corollary 9 and K_Z is fixed, positive and reversible, with absolute nonconstant spectral radius at most ρ. Under the same parameter conditions, with $B = 1/\min \pi_X$ for the full product, the randomized KL order is $\Theta_{K,\tau}(t/L)$ (Lemma 18). The following consequence separates the full deterministic class from this hidden-phase sampler in total variation throughout that finite-state class; equivalently, for each deterministic policy there is a bounded output test witnessing the separation.

Corollary 11 (Bounded output discrimination)**.** *Under the hypotheses of Theorem 10, there is $a_K > 0$ such that, for all $R, s \geq 1$,*

$$\inf_{\sigma \in \mathcal{D}_R} \mathrm{TV}(P_n, Q_\sigma) \geq 1 - e^{-a_K t}.$$

When $0 < t \leq \tau$ and $B\rho^L \leq 1/2$, the hidden-phase sampler satisfies $\mathrm{TV}(P_n, \overline{Q}) = O_{K,\tau}(\sqrt{t/L})$. Thus, along $L \to \infty$ and $t \to t_0 > 0$, the deterministic infimum stays positive while the hidden-phase distance tends to zero.

Here $\mathrm{TV}(P, Q) = \sup_{0 \leq f \leq 1} |\mathbb{E}_P f - \mathbb{E}_Q f|$. The distinguishing test may depend on the deterministic policy; the bound is uniform over those policies. The proof contracts conditional Hellinger affinities along the adaptive policy tree, then uses a pathwise dependence witness in every cell. Appendix D.3 gives the proof and matching squared-Hellinger rates for both policy classes on positive-correlation binary chains.

D.2 Positive reversible chains: complete proof

D.2.1 Model and theorem

This appendix proves Theorem 10; Theorem 12 below restates it with explicit constants. Let $\mathcal{X}$ be a fixed finite alphabet of size $q \geq 2$. Fix a stochastic matrix K with $K(x, y) > 0$ for all x, y, reversible with respect to its stationary distribution $\pi > 0$. Let $\Pi(x, y) = \pi(y)$, assume $K \neq \Pi$, and set

$$\rho = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } K \text{ on } \mathbf{1}^\perp\} \in (0, 1).$$

The orthogonal complement here is in $L^2(\pi)$. Reversibility makes K self-adjoint, so these eigenvalues are real. Strict positivity makes the eigenvalue one simple and gives $|\lambda| < 1$ on $\mathbf{1}^\perp$: an eigenfunction attaining maximal absolute value cannot preserve that value under a strictly positive average unless it is constant. If all nonconstant eigenvalues were zero, the spectral decomposition would give $K = \Pi$. These facts justify $0 < \rho < 1$, including when the dominant nonconstant eigenvalues are negative or repeated.

The stationary target law is

$$P_n(x_1, \ldots, x_n) = \pi(x_1) \prod_{j=1}^{n-1} K(x_j, x_{j+1}).$$

Policies and the classes $\mathcal{D}_R, \mathcal{R}_R$ are exactly those of Section 3; no additional restriction on value adaptation is imposed. We retain the pathwise round cap and the independent, hidden scheduling tape.

Write

$$B = \frac{1}{\min_x \pi(x)}, \qquad \delta = \min\left\{1, \min_{x,y} \frac{K(x, y)}{\pi(y)}\right\}, \qquad U = \max\left\{1, \max_{x,y} \frac{K(x, y)}{\pi(y)}\right\}. \tag{43}$$

Choose a real eigenfunction f with eigenvalue θ, $|\theta| = \rho$, normalized by $\mathbb{E}_\pi f = 0$, $\mathbb{E}_\pi f^2 = 1$. Set

$$F = \|f\|_\infty, \qquad c_0 = \frac{\delta^4}{72U^4F^4}, \qquad c_K = \frac{c_0}{2} = \frac{\delta^4}{144U^4F^4} > 0. \tag{44}$$

Any one such f may be fixed; $1 \leq F^2 \leq B$. In particular $\delta^4/(144U^4B^2)$ is an alternative constant independent of this choice.

Theorem 12 (Deterministic order and hidden-phase separation)**.** *Let K, π, ρ be fixed as above. For integers $R, s \geq 1$, set*

$$L = 2^{R-1}, \qquad n = (s+1)L, \qquad u = \rho^L, \qquad t = su^2.$$

For every such R, s,

$$c_K t \;\leq\; \inf_{\sigma \in \mathcal{D}_R} D_{\mathrm{KL}}(P_n \| Q_\sigma) \;\leq\; Bt. \tag{45}$$

For each fixed $0 < \tau < \infty$, there are finite constants $C_{K,\tau}, C'_{K,\tau}$, independent of R, s, n, such that whenever $Bu \leq 1/2$ and $0 < t \leq \tau$, the following R-round hidden-phase policy has output law $\overline{Q}$ satisfying

$$V := \chi^2(\overline{Q} \| P_n) \leq \frac{C_{K,\tau} t}{L}, \tag{46}$$

$$D_{\mathrm{KL}}(P_n \| \overline{Q}) \leq (2 + 4Bt)V + 4B\sqrt{D_K t V / L}, \qquad D_K = 1 + \frac{10B\rho + 2}{1 - \rho}. \tag{47}$$

In particular, $D_{\mathrm{KL}}(P_n \| \overline{Q}) \leq C'_{K,\tau} t / L$. The policy chooses I uniformly from $\{1, \ldots, L\}$, independently samples values at the $s+1$ first-round positions $I + jL$, $0 \leq j \leq s$, from π, and completes all remaining intervals by balanced exact singleton separators in at most $R - 1$ further rounds. It returns only the sequence.

Consequently, for fixed $t_0 > 0$, $R \to \infty$ and $s = \lfloor t_0 / \rho^{2L} \rfloor$, eventually $s \geq 1$, $t \to t_0$, the hidden-phase KL is $O_{K,t_0}(L^{-1}) \to 0$, and the deterministic optimum has lower limit at least $c_K t_0 > 0$.

The theorem allows an arbitrary number of reversible correlation modes. For the upper bound, Appendices D.2.4–D.2.6 replace the binary transfer calculation by spectral and boundary estimates; Appendix D.2.7 converts density variance to KL. For the lower bound, only the boundary-uniform pair-information estimate (Appendix D.2.8) is new: the history identity and charging rule are those of Appendices B.2.5 and B.2.8. These shared arguments use the conditional-total-correlation decomposition of Wen et al. (2026) and the Markov-bisection principle of Bansal et al. (2026).

D.2.2 Density coordinates and normalized lag powers

For a density kernel $A(x, y)$, write

$$(Ag)(x) = \sum_y \pi(y) A(x, y) g(y), \quad (AC)(x, y) = \sum_z \pi(z) A(x, z) C(z, y), \quad A^*(x, y) = A(y, x)$$

for real kernels. A centered kernel has zero π-weighted row and column sums. The density kernel identically equal to one represents the projection onto constants. The Hilbert–Schmidt inner product is

$$\langle A, C \rangle_{\mathrm{HS}} = \sum_{x,y} \pi(x)\pi(y) A(x, y) C(x, y) = \mathrm{tr}(A^* C).$$

Thus transposition in density coordinates represents the $L^2(\pi)$ adjoint, while every operator product includes a π weight. Traces may equivalently be computed after conjugation to the orthonormal coordinate basis $\mathbf{1}_{\{x\}} / \sqrt{\pi(x)}$.

For $d \geq 1$ define

$$S_d(x, y) = \frac{K^d(x, y)/\pi(y) - 1}{\rho^d}. \tag{48}$$

The spectral decomposition of K gives, for a real orthonormal eigenbasis $f_1 = \mathbf{1}, f_2, \ldots, f_q$,

$$S_d(x, y) = \sum_{j=2}^{q} \left(\frac{\lambda_j}{\rho} \right)^d f_j(x) f_j(y).$$

Hence S_d is centered, self-adjoint, and has $L^2(\pi)$ operator norm at most one. All S_d commute, and

$$S_h S_k = S_{h+k} \qquad (h, k \geq 1). \tag{49}$$

Indeed the centered part of K^d annihilates constants, so multiplication of centered parts agrees with the centered part of K^{h+k} after the factors $\rho^h \rho^k$ are removed.

Any self-adjoint contraction A in density coordinates obeys

$$|A(x,y)| \le \frac{1}{\sqrt{\pi(x)\pi(y)}} \le B, \qquad \sum_y \pi(y)A(x,y)^2 \le \frac{1}{\pi(x)} \le B. \tag{50}$$

To see this, its ordinary matrix in the orthonormal coordinate basis has entries $\sqrt{\pi(x)\pi(y)}A(x,y)$ and operator norm at most one, so every entry and every row Euclidean norm are at most one. These bounds apply to S_d and to products of commuting self-adjoint contractions.

D.2.3 Legal phase policies, densities, and deterministic upper bound

The separator argument in Appendix B.2.1 uses only the Markov property, so it applies to this alphabet as well. For phase i, the interior gaps have $L-1$ sites and the tails have $i-1$ and $L-i$ sites; exact completion therefore takes at most $R-1$ additional rounds. For $R=1$, $L=1$ and the first batch already contains all sites.

Under the target law, the values at the first-round positions form a stationary chain with kernel K^L. Put $E_L = S_L$. Exact conditional completion gives the full output density ratio

$$T_i := \frac{Q_i}{P_n} = \prod_{j=0}^{s-1} \frac{1}{1+uE_L(X_{i+jL}, X_{i+(j+1)L})}, \qquad T = \frac{1}{L}\sum_{i=1}^{L} T_i, \qquad \mathbb{E}_{P_n} T_i = \mathbb{E}_{P_n} T = 1. \tag{51}$$

Stationarity then gives

$$D_{\mathrm{KL}}(P_n \| Q_i) = sI_L, \tag{52}$$
$$I_L = \sum_x \pi(x) D_{\mathrm{KL}}(K^L(x,\cdot)\|\pi) \le u^2 \sum_{x,y} \pi(x)\pi(y)E_L(x,y)^2 \le Bu^2.$$

The final inequality uses the row bound in (50) for E_L, followed by averaging over x: $\sum_{x,y}\pi(x)\pi(y)E_L(x,y)^2 \le B$. Here $D_{\mathrm{KL}} \le \chi^2$ follows from $\log z \le z-1$, and all rows of K^L are strictly positive. No small-u assumption is used in (52). Each fixed phase is a deterministic legal policy, proving the upper bound in (45) for all R, s.

Assume now $Bu \le 1/2$. Define the ordinary nonnegative matrix

$$H(x,y) = \frac{\pi(y)}{1+uE_L(x,y)}.$$

Integrating the target chain on the first-round positions in (51) gives $\mathbb{E}T_i^2 = \pi H^s \mathbf{1}$. Centering and (50) imply

$$(H\mathbf{1})(x) = 1 + u^2 \sum_y \pi(y) \frac{E_L(x,y)^2}{1+uE_L(x,y)} \le 1 + 2Bu^2,$$

using $(1+w)^{-1} = 1 - w + w^2/(1+w)$. Thus

$$\mathbb{E}T_i^2 \le e^{2Bt}, \qquad 0 \le \mathbb{E}T_iT_j \le e^{2Bt}, \tag{53}$$

the latter by Cauchy–Schwarz.

D.2.4 A uniform analytic transfer lemma

Let $\mathfrak{C}_\pi$ consist of pairs (C,D) of real, centered, commuting, self-adjoint contractions on $L^2(\pi)$. In the finite-dimensional coordinate space this set is closed and bounded, hence compact. For such a pair set $E = CD = DC$. It too is centered, self-adjoint and a contraction. For complex parameters a, b near zero define an ordinary matrix on $\mathcal{X}^2$ and a row vector by

$$W(a,b)_{(x,y),(p,z)} = \frac{\pi(p)\pi(z)(1+bD(y,p))(1+aC(p,z))}{(1+abE(x,p))(1+abE(y,z))}, \tag{54}$$
$$\mu_a(x,y) = \pi(x)\pi(y)(1+aC(x,y)).$$

Let P_0 be the projection $(P_0 g)(x,y) = \sum_{p,z}\pi(p)\pi(z)g(p,z)$. All matrix norms below are the induced maximum absolute row-sum norm, denoted $\|\cdot\|_\infty$.

Lemma 13 (Uniform analytic cancellation)**.** *There exist $\epsilon > 0$ and $A \geq 1$ depending only on π (and its fixed finite alphabet), common to all $(C, D) \in \mathfrak{C}_\pi$, with the following properties. On $0 \leq a, b \leq \epsilon$, W has a simple real eigenvalue $\lambda > 3/4$ and a rank-one commuting spectral projection P such that*

$$\lambda - 1 = a^2 b^2 H_{C,D}(a, b), \qquad H_{C,D}(0, 0) = 0, \tag{55}$$

$$|\lambda - 1| \leq A a^2 b^2 (a + b), \tag{56}$$

$$|\mu_a P\mathbf{1} - 1| \leq A(a + b), \qquad N := W - \lambda P,$$

$$PN = NP = 0, \qquad \|N\|_\infty \leq A(a + b) \leq 1/2. \tag{57}$$

The eigenvalue, projection, and $H_{C,D}$ are holomorphic in (a, b) on a common complex neighborhood of zero; their bounds are uniform over $\mathfrak{C}_\pi$. One admissible choice is

$$r_0 = \frac{1}{1024B}, \qquad A = r_0^{-5} + 20 r_0^{-1} + 1, \qquad \epsilon = \min\left\{ \frac{r_0}{2}, \frac{1}{4A}, \frac{1}{2B} \right\}. \tag{58}$$

Proof. Common holomorphic domain and spectral contour. For $|a|, |b| \leq r_0$, (50) gives $|abE(x, p)| \leq B r_0^2 < 1$. Thus all entries of W are holomorphic on a neighborhood of this closed polydisc, uniformly over the compact parameter family. Subtracting the denominator product from the numerator product in (54), and using that the $\pi(p)\pi(z)$ weights sum to one, gives

$$\|W - P_0\|_\infty \leq \frac{2Br_0 + B^2 r_0^2 + 2Br_0^2 + B^2 r_0^4}{(1 - Br_0^2)^2} \leq 8Br_0 = \frac{1}{128}.$$

For example, the last elementary bound follows from $B \geq 1$, $r_0 \leq 1$, and $Br_0 \leq 1/1024$.

On $\Gamma = \{z : |z - 1| = 1/4\}$,

$$(zI - P_0)^{-1} = \frac{P_0}{z - 1} + \frac{I - P_0}{z}, \qquad \left\|(zI - P_0)^{-1}\right\|_\infty \leq 4 + \frac{8}{3} < 7.$$

The Neumann series for $(zI - W)^{-1}$ therefore converges uniformly on Γ and the common polydisc, since $7/128 < 1$. Define

$$P = \frac{1}{2\pi \mathrm{i}} \int_\Gamma (zI - W)^{-1}\, dz.$$

This projection is holomorphic in a, b, continuous in (C, D), and commutes with W. Along $P_0 + v(W - P_0)$, $0 \leq v \leq 1$, the same resolvent bound holds, so its rank is the constant rank one of the initial contour projection. Hence there is exactly one simple eigenvalue in the disk $|z - 1| < 1/4$, given holomorphically by $\lambda = \mathrm{tr}(WP)$; in particular $WP = PW = \lambda P$. For real a, b, complex conjugation and uniqueness imply that λ is real, so $\lambda > 3/4$.

The resolvent identity gives, uniformly on the closed polydisc,

$$\|P - P_0\|_\infty \leq \frac{1}{4} \frac{49/128}{1 - 7/128} = \frac{49}{484} < \frac{1}{8}.$$

Consequently $\|P\|_\infty < 2$, $\|W\|_\infty < 2$, $|\lambda| < 5/4$, and $\|N\|_\infty < 5$. Also $\|\mu_a\|_1 \leq 1 + B|a| < 2$, so

$$|\gamma(a, b)| < 5, \qquad \gamma := \mu_a P\mathbf{1} - 1.$$

At the origin $W = P_0$, $P = P_0$, $\lambda = 1$, $N = 0$, and $\gamma = 0$. Moreover $\|P - P_0\|_\infty < 1/8$ ensures that the denominator in

$$r = \frac{P\mathbf{1}}{\langle \pi \otimes \pi, P\mathbf{1} \rangle}$$

is nonzero. This is a holomorphic right eigenvector satisfying $P_0 r = \mathbf{1}$, uniformly available for the Taylor calculation below.

Axis identities and divisibility. At $b = 0$, every row of $W(a, 0)$ equals μ_a, with both marginals π. Thus $W(a, 0)$ is a rank-one projection with eigenvalue one. The derivative in b of the undivided numerator row sum is zero, and the reciprocal denominator contributes

$$-a \sum_{p,z} \mu_a(p, z)\{E(x, p) + E(y, z)\} = 0$$

by the marginals and centering. Therefore $\partial_b(W\mathbf{1})(a,0)=0$. For a left eigenvector ℓ normalized by $\ell r=1$, differentiation of $Wr=\lambda r$ gives $\partial_b\lambda=\ell(\partial_b W)r$. On this axis $r=\mathbf{1}$ and $\ell=\mu_a$, so

$$\lambda(a,0)=1, \qquad \partial_b\lambda(a,0)=0.$$

At $a=0$,

$$W(0,b)_{(x,y),(p,z)}=\pi(p)\pi(z)(1+bD(y,p)).$$

It preserves constants, has stationary row vector $\pi\otimes\pi$, and its square is P_0, all by the zero weighted row and column sums of D. Differentiation gives exactly

$$\partial_a(W\mathbf{1})(x,y)|_{a=0}=-b^2\sum_p \pi(p)E(x,p)D(y,p)=-b^2(ED^*)(x,y). \tag{59}$$

Its mean under $\pi\otimes\pi$ is zero. Hence $\lambda(0,b)=1$ and $\partial_a\lambda(0,b)=0$. These identities hold on open axis disks, so the holomorphic Taylor series of $\lambda-1$ has a-degree and b-degree both at least two. This proves the factorization in (55) apart from $H_{C,D}(0,0)=0$.

Fourth-order cancellation in weighted coordinates. Below every expectation over (p,z) uses $\pi(p)\pi(z)$ independently, and products of operators use their weighted convention. Write $G=ED^*$. The numerator in (54), after removing $\pi(p)\pi(z)$, is

$$1+aC(p,z)+bD(y,p)+abD(y,p)C(p,z),$$

and its row mean is exactly one. The reciprocal denominator through total degree four is

$$1-ab\{E(x,p)+E(y,z)\}+a^2b^2\{E(x,p)^2+E(y,z)^2+E(x,p)E(y,z)\}.$$

The omitted reciprocal terms have total degree at least six. In the product, all degree-two terms vanish by centering. Among the degree-three terms, only the product of $bD(y,p)$ with $-abE(x,p)$ survives; it equals $-ab^2G(x,y)$. At degree four, the squared terms give

$$a^2b^2\left\{\sum_p \pi(p)E(x,p)^2+\sum_z \pi(z)E(y,z)^2\right\}.$$

The cross term $E(x,p)E(y,z)$ averages to zero. Multiplying $abD(y,p)C(p,z)$ by $-abE(x,p)$ gives zero because the row mean of C is zero. Multiplying it by $-abE(y,z)$ gives

$$-a^2b^2\sum_z \pi(z)E(y,z)(DC)(y,z)=-a^2b^2\sum_z \pi(z)E(y,z)^2,$$

where commutation is used in $DC=E$. Thus, uniformly over the compact family and over current states,

$$(W\mathbf{1})(x,y)=1-ab^2G(x,y)+a^2b^2\sum_p \pi(p)E(x,p)^2+O((|a|+|b|)^5). \tag{60}$$

Uniformity of the remainder also follows from the already established common holomorphic domain and uniform bounds.

The degree-one operator in W is $a\mathcal{A}+b\mathcal{B}$, where

$$(\mathcal{A}g)(x,y)=\langle C,g\rangle_{\mathrm{HS}}, \qquad (\mathcal{B}g)(x,y)=\sum_{p,z}\pi(p)\pi(z)D(y,p)g(p,z).$$

Expand the normalized eigenvector as $r=\mathbf{1}+r_1+r_2+\cdots$ and the eigenvalue as $\lambda=1+\lambda_1+\lambda_2+\cdots$, with homogeneous terms of total degree and $P_0r_k=0$. Comparing degrees in $Wr=\lambda r$, using (60), gives $r_1=r_2=0$, $\lambda_1=\lambda_2=\lambda_3=0$, and $r_3=-ab^2G$. Here $P_0G=0$ by centering. The row means of G vanish, so $\mathcal{B}G=0$. The degree-four eigenvalue is therefore

$$\lambda_4=a^2b^2\{\|E\|_{\mathrm{HS}}^2-\langle C,ED^*\rangle_{\mathrm{HS}}\}=0,$$

since cyclicity of the finite-dimensional trace gives

$$\langle C,ED^*\rangle_{\mathrm{HS}}=\mathrm{tr}(C^*ED^*)=\mathrm{tr}(D^*C^*E)=\mathrm{tr}((CD)^*E)=\|E\|_{\mathrm{HS}}^2. \tag{61}$$

This proves $H_{C,D}(0,0)=0$. The trace step uses the adjoint in the stated weighted coordinates; the row-sum step is where the relation $DC=CD$ is essential.

Uniform derivative bounds and concrete constants. Write $\lambda - 1 = \sum_{i,j\geq 0} c_{ij} a^i b^j$. The contour locates $|\lambda - 1| < 1/4$ on the common polydisc, so Cauchy's estimate gives

$$|c_{ij}| \leq \frac{1}{4r_0^{i+j}}.$$

The axis identities and cancellation show $c_{ij} = 0$ when $i < 2$ or $j < 2$, and $c_{22} = 0$. For $|a|, |b| \leq r_0/2$, put $x = |a|/r_0$, $y = |b|/r_0$. Summing the remaining coefficients yields

$$\begin{aligned} |H_{C,D}(a,b)| &\leq \frac{1}{4r_0^4}\left\{\frac{1}{(1-x)(1-y)} - 1\right\} \\ &\leq \frac{|a|+|b|}{r_0^5}, \end{aligned}$$

because the expression in braces is at most $4(x+y)$. Matrix-valued Cauchy estimates applied to N, and scalar estimates applied to γ, with their uniform bounds 5 and zero constant terms, similarly give

$$\|N\|_\infty \leq \frac{20}{r_0}(|a|+|b|), \qquad |\gamma| \leq \frac{20}{r_0}(|a|+|b|).$$

For the matrix estimate, integrate coefficients as matrices and use the induced norm before summing; no dimension-dependent entrywise accumulation is needed. The choices in (58) now imply every assertion of the lemma, including $2A\epsilon \leq 1/2$. For $0 \leq a \leq \epsilon$, μ_a is a probability row vector, since it sums to one and $|aC(x,y)| \leq 1/2$. All constants are common to the entire compact family, including normalized lag powers that vary with L and h. □

D.2.5 Finite boundary and transient cancellation

The finite initial and terminal factors enjoy the same cancellation as the leading eigenvalue.

Lemma 14 (Boundary and transient cancellation)**.** *Use the notation and constants of Lemma 13, and put*

$$\gamma = \mu_a P\mathbf{1} - 1, \qquad \eta_s = \mu_a N^s \mathbf{1} \quad (s \geq 1).$$

There are holomorphic functions g, j_s near zero such that

$$\gamma = a^2b^2 g(a,b), \qquad \eta_s = a^2b^2 j_s(a,b), \qquad g(0,0) = j_s(0,0) = 0. \tag{62}$$

Uniformly over $(C,D) \in \mathfrak{C}_\pi$, all integers $s \geq 1$, and complex $|a|, |b| \leq \epsilon/2$,

$$|\gamma| \leq 20\epsilon^{-5}|a|^2|b|^2(|a|+|b|), \tag{63}$$

$$|\eta_s| \leq 2^{3-s}\epsilon^{-5}|a|^2|b|^2(|a|+|b|). \tag{64}$$

Proof. The complex Cauchy estimates in Lemma 13 give $\|N\|_\infty \leq A(|a|+|b|) \leq 1/2$ on the closed ϵ-polydisc. On that polydisc, $|\gamma| \leq 5$ and $\|\mu_a\|_1 \leq 2$, hence

$$|\eta_s| \leq 2^{1-s}. \tag{65}$$

In particular, η_s and its first derivatives tend locally uniformly to zero as $s \to \infty$.

Write $F_s = \mu_a W^s \mathbf{1}$ and $\nu = \pi \otimes \pi$. We first show, for each $s \geq 1$, the exact axis identities

$$F_s(a,0) = F_s(0,b) = 1, \qquad \partial_b F_s(a,0) = \partial_a F_s(0,b) = 0. \tag{66}$$

At $b = 0$, W has every row equal to μ_a, preserves constants, and $(\partial_b W)\mathbf{1} = 0$, as established in the proof of Lemma 13. Differentiating $W^s\mathbf{1}$ proves the first transverse identity. At $a = 0$, put $W_b = W(0,b)$. Its centered kernel gives $W_b\mathbf{1} = \mathbf{1}$ and $\nu W_b = \nu$, while (59) gives

$$v := (\partial_a W)(0,b)\mathbf{1} = -b^2 G, \qquad G = ED^*, \qquad \nu G = 0.$$

Since $(\partial_a \mu_a)\mathbf{1} = 0$, differentiation yields

$$\partial_a F_s(0,b) = \sum_{k=0}^{s-1} \nu W_b^k v = s\nu v = 0.$$

All these identities hold on complex axis disks.

Next, $F_s - 1$ has no Taylor terms of total degree at most four, for every $s \geq 1$. To see this with the finite boundaries retained, write (60) as

$$W\mathbf{1} = \mathbf{1} + v_3 + v_4 + O_5, \qquad v_3 = -ab^2 G, \qquad v_4(x,y) = a^2b^2 \sum_p \pi(p) E(x,p)^2,$$

where O_5 denotes terms of total degree at least five. Put $\kappa = \|E\|_{\mathrm{HS}}^2$, and let $W_{[1]} = a\mathcal{A} + b\mathcal{B}$ be the degree-one operator from the preceding proof. Centering and (61) give

$$P_0 v_3 = 0, \qquad P_0 v_4 = a^2b^2\kappa\mathbf{1}, \qquad \mathcal{B}G = 0, \qquad W_{[1]} v_3 = -a^2b^2\kappa\mathbf{1}.$$

Thus induction, multiplying by W and keeping degrees at most four, gives $W^s\mathbf{1} = \mathbf{1} + v_3 + v_4 + O_5$ for each $s \geq 1$. The initial row is $\mu_a(x,y) = \nu(x,y)(1 + aC(x,y))$, so

$$F_s = 1 + a^2b^2\{\kappa - \langle C, G\rangle_{\mathrm{HS}}\} + O_5 = 1 + O_5.$$

Only Taylor coefficients are being compared here; no bound uniform in s on this formal remainder is needed.

The exact spectral decomposition gives

$$F_s = \lambda^s(1+\gamma) + \eta_s. \tag{67}$$

On either axis, $\lambda = 1$ and its transverse first derivative is zero. Restricting (67) to that axis and using (66), then letting $s \to \infty$, shows that γ vanishes there. Differentiating transversely and using local uniform convergence of the derivatives of η_s shows that the corresponding derivative of γ also vanishes. Hence γ is divisible by a^2b^2. To determine its coefficient of a^2b^2, take $s = 5$ in (67): $N(0,0) = 0$ implies that η_5 has no terms below degree five, and the same is true of $\lambda^5 - 1$ by (55). Since $F_5 = 1 + O_5$, this coefficient of γ is zero. Subtracting $\lambda^s(1+\gamma)$ from F_s now proves the same axis and fourth-degree cancellations for η_s for every $s \geq 1$, including $s = 1$. This proves (62).

Finally, if a holomorphic scalar function f bounded by M on the ϵ-polydisc has these cancellations, Cauchy's coefficient bound, with $x = |a|/\epsilon \leq 1/2$ and $y = |b|/\epsilon \leq 1/2$, gives

$$|f(a,b)| \leq M\epsilon^{-4}|a|^2|b|^2 \left\{ \frac{1}{(1-x)(1-y)} - 1 \right\} \leq 4M\epsilon^{-5}|a|^2|b|^2(|a| + |b|).$$

Applying this to γ with $M = 5$ and to η_s with $M = 2^{1-s}$ proves the stated uniform bounds. □

D.2.6 Exact cross moments and uniform long powers

For a lag $1 \leq h < L$ and phases $i, i+h \in \{1, \ldots, L\}$, put

$$a = \rho^h, \quad b = \rho^{L-h}, \quad C = S_h, \quad D = S_{L-h}, \quad E = CD = S_L.$$

These operators belong to $\mathfrak{C}_\pi$ and $ab = u$. The two sets of first-round values interleave as $x_0, y_0, x_1, y_1, \ldots, x_s, y_s$, with gaps $h, L-h, h, \ldots, h$. Marginalizing sites outside the first and last interleaved first-round positions, using stationarity and stochasticity, gives the exact identity

$$c_s(h) := \mathbb{E}_{P_n}[T_i T_{i+h}] = \mu_a W(a,b)^s \mathbf{1}. \tag{68}$$

There are $s+1$ state pairs and s transfers. The initial law is μ_a and the terminal function is $\mathbf{1}$; these include both boundaries of the finite nonperiodic path.

If $a, b \leq \epsilon/2$, Lemmas 13 and 14 give

$$c_s(h) = \lambda^s(1+\gamma) + \eta_s, \qquad |\lambda^s - 1| \leq s|\lambda - 1|e^{s|\lambda-1|} \leq At(a+b)e^{A\tau\epsilon}.$$

Here $sa^2b^2 = t \leq \tau$, $a + b \leq \epsilon$, and $\lambda > 0$; also $|\lambda|^s \leq e^{A\tau\epsilon}$. Since $a^2b^2 = t/s \leq t$ and $2^{3-s} \leq 4$, the exact finite-boundary bounds yield

$$|c_s(h) - 1| \leq C_\tau^{\mathrm{int}} t(a+b), \qquad C_\tau^{\mathrm{int}} = (A + 20\epsilon^{-5})e^{A\tau\epsilon} + 4\epsilon^{-5}. \tag{69}$$

This estimate is uniform over all $s \geq 1$, including $s = 1$.

For the remaining lags, center before applying Cauchy–Schwarz. Since $\mathbb{E}T_i = 1$, (53) gives, for all phases i, j,

$$|\mathbb{E}T_iT_j - 1| = |\mathbb{E}[(T_i - 1)(T_j - 1)]| \le \sqrt{\mathrm{Var}(T_i)\,\mathrm{Var}(T_j)} \le e^{2Bt} - 1 \le d_\tau t, \qquad d_\tau = 2Be^{2B\tau}. \tag{70}$$

Choose a fixed integer $m \ge 1$ with $\rho^m \le \epsilon/2$. At most $2m$ lags in $1, \dots, L-1$ fail $a, b \le \epsilon/2$; bound them, and the diagonal, by (70). Stationarity makes all diagonal moments equal and each cross moment depend only on h, so

$$V = \frac{\mathbb{E}T_1^2 - 1}{L} + \frac{2}{L^2}\sum_{h=1}^{L-1}(L-h)(c_s(h) - 1). \tag{71}$$

Use the absolute bounds above, $L - h \le L$, and

$$\sum_{h=1}^{L-1}(\rho^h + \rho^{L-h}) \le \frac{2\rho}{1-\rho}$$

to obtain (46) with the explicit finite choice

$$C_{K,\tau} = (1 + 4m)d_\tau + \frac{4C_\tau^{\mathrm{int}}\rho}{1-\rho}. \tag{72}$$

Here A, ϵ depend only on π through (58); m also depends on ρ; and $d_\tau, C_\tau^{\mathrm{int}}$ also depend on τ. No constant depends on s, L or a positive lower bound on t. The exceptional-lag overcount remains valid for $L < 2m$, and the lag sum is empty for $L = 1$. No sign condition on individual cross covariances is used.

D.2.7 From density variance to target-to-output KL

The density estimate alone does not control the low-density tail of forward KL. The additional input is concentration of the average phase log cost. Although a phase spans the full chain, this average is a sum of observables that each depend on only two endpoints.

Lemma 15 (Two-endpoint covariance bound)**.** *Let (X_j) be the stationary chain with kernel K. For any $L, N \ge 1$ and real function g with $|g| \le G$, put $Y_j = g(X_j, X_{j+L})$. Then*

$$\mathrm{Var}\left(\sum_{j=1}^{N} Y_j\right) \le D_K N G^2, \qquad D_K = 1 + \frac{10B\rho + 2}{1-\rho}. \tag{73}$$

The constant is independent of the separation L.

Proof. The lag-density bound (50) gives $\delta_d := \max_x \sum_y |K^d(x,y) - \pi(y)| \le B\rho^d$. For $1 \le h < L$, the four ordered endpoints have joint law

$$J(a,b,c,d) = \pi(a)K^h(a,b)K^{L-h}(b,c)K^h(c,d).$$

Replacing its three transitions by the stationary kernel gives an ℓ^1 distance at most $2\delta_h + \delta_{L-h}$ from $\pi^{\otimes 4}$. The product of the two endpoint-pair marginals, $J_{\mathrm{ind}}(a,b,c,d) = \pi(a)K^L(a,c)\pi(b)K^L(b,d)$, is at distance at most $2\delta_L$ from the same law. Hence

$$|\,\mathrm{Cov}(Y_1, Y_{1+h})| \le BG^2(2\rho^h + \rho^{L-h} + 2\rho^L), \qquad 1 \le h < L. \tag{74}$$

For $h \ge L$, write $m = \mathbb{E}Y_1$ and define $f(x) = \mathbb{E}[Y_1 \mid X_{1+L} = x]$ and $b(x) = \mathbb{E}[Y_1 \mid X_1 = x]$. Conditional independence across the intermediate endpoint gives

$$\mathrm{Cov}(Y_1, Y_{1+h}) = \langle f - m, K^{h-L}(b - m)\rangle_{L^2(\pi)}.$$

This formula includes $h = L$, when the pairs share one endpoint. Reversibility bounds the centered operator norm by ρ^{h-L}, and conditional expectation contracts variance. Thus

$$|\,\mathrm{Cov}(Y_1, Y_{1+h})| \le G^2\rho^{h-L}, \qquad h \ge L.$$

The short-lag bound sums to at most $5BG^2\rho/(1-\rho)$, using $(L-1)\rho^L \le \sum_{h=1}^{L-1}\rho^h$. The remaining lags sum to at most $G^2/(1-\rho)$. In the stationary variance identity, bound $N - h$ by N and $\mathrm{Var}(Y_1)$ by G^2 to obtain (73). The same argument covers $L = 1$ and $N = 1$ with empty sums omitted. □

Set $g_L(x,y) = \log[K^L(x,y)/\pi(y)]$ and $A_{\log} = L^{-1}\sum_{i=1}^{L}(-\log T_i)$. The phase density identity reindexes exactly as

$$A_{\log} = \frac{1}{L}\sum_{k=1}^{sL} g_L(X_k, X_{k+L}).$$

All endpoints lie in the length-$n = (s+1)L$ target path. Since $Bu \le 1/2$, $|g_L| \le 2Bu$ by (50). Lemma 15 and (52) therefore give

$$\mathrm{Var}(A_{\log}) \le \frac{4B^2 D_K t}{L}, \qquad 0 \le \mathbb{E}A_{\log} = sI_L \le Bt. \tag{75}$$

AM–GM yields $-\log T \le A_{\log}$. The centered lower-tail argument (20) uses only this domination, $\mathbb{E}T = 1$, and $V = \mathrm{Var}(T)$; it is independent of the alphabet. Substituting (75) gives

$$D_{\mathrm{KL}}(P_n\|\overline{Q}) \le (2 + 4\mathbb{E}A_{\log})V + 2\sqrt{\mathrm{Var}(A_{\log})V} \le (2+4Bt)V + 4B\sqrt{D_K t V/L}.$$

This proves (47), including $V = 0$. Combining it with (46) gives

$$D_{\mathrm{KL}}(P_n\|\overline{Q}) \le \frac{C'_{K,\tau}t}{L}, \qquad C'_{K,\tau} = (2+4B\tau)C_{K,\tau} + 4B\sqrt{D_K C_{K,\tau}}.$$

The estimate is uniform over $0 < t \le \tau$, including $s = 1$; it does not require a positive lower bound on t. A matching randomized converse throughout the general reversible class remains a separate question.

D.2.8 Uniform conditional information by inverse tilting

Lemma 16 (Lipschitz normalized tilting)**.** *Let A_1, A_2 be probability tables on a finite set and let w have entries in $[w_-, w_+]$, with $0 < w_- \le w_+ < \infty$. For*

$$\mathcal{T}_w(A) = \frac{wA}{\langle w, A\rangle}$$

(entrywise multiplication), one has

$$\|\mathcal{T}_w(A_1) - \mathcal{T}_w(A_2)\|_1 \le 2\frac{w_+}{w_-}\|A_1 - A_2\|_1.$$

If the set is a Cartesian product and $w(x,y) = v(x)z(y)$, tilting a product probability table gives a product probability table.

Proof. Let $Z_i = \langle w, A_i\rangle \in [w_-, w_+]$. Split the difference into $w(A_1 - A_2)/Z_1 + wA_2(1/Z_1 - 1/Z_2)$. The first term has ℓ^1 norm at most $(w_+/w_-)\|A_1 - A_2\|_1$. The norm of the second equals $|Z_1 - Z_2|/Z_1$, with the same bound. For a product table and product weights, both the numerator and the normalizer factor into one-coordinate terms, proving the last assertion. □

Lemma 17 (Boundary-uniform pair information)**.** *Suppose $v < w$ are unrevealed sites at distance $d = w - v \le L$, with no previously revealed site strictly between them. Conditional on any revealed assignment, their mutual information is at least $c_0\rho^{2L}$, with c_0 from* (44)*. Other unrevealed sites selected in the same batch are not conditioned on here and are marginalized out.*

Proof. First, for every $k \ge 1$,

$$\delta \le \frac{K^k(x,y)}{\pi(y)} \le U. \tag{76}$$

Indeed $K^k(x,y)/\pi(y)$ averages $K(z,y)/\pi(y)$ against the probability row $K^{k-1}(x,z)$, including $k = 1$ with $K^0 = I$.

Let $J_d(x,y) = \pi(x)K^d(x,y)$, whose two marginals are π. The fixed eigenfunction in (44) satisfies

$$\sum_{x,y} f(x)f(y)\{J_d(x,y) - \pi(x)\pi(y)\} = \langle f, K^d f\rangle_\pi = \theta^d.$$

Taking absolute values and using $|f(x)f(y)| \le F^2$ gives

$$\|J_d - \pi \otimes \pi\|_1 \ge \frac{\rho^d}{F^2}. \tag{77}$$

The absolute spectral bound handles either sign of the dominant eigenvalue.

By the path Markov factorization, only the nearest revealed boundary on each side affects the conditional table of the pair. A left boundary z_ℓ at distance $k \ge 1$ gives likelihood

$$\ell_x = \frac{K^k(z_\ell, x)}{\pi(x)} = \frac{K^k(x, z_\ell)}{\pi(z_\ell)},$$

where the second equality follows from reversibility of K^k. A right boundary z_r at distance $j \ge 1$ gives $r_y = K^j(y, z_r)/\pi(z_r)$. If either boundary is absent its likelihood factor is one. All remaining revealed evidence multiplies by factors independent of the pair and cancels upon normalization. Thus its exact conditional table is

$$M(x,y) = \frac{\ell_x r_y J_d(x,y)}{Z}, \qquad Z = \sum_{x,y} \ell_x r_y J_d(x,y), \qquad \ell_x, r_y \in [\delta, U]. \tag{78}$$

In particular $Z \in [\delta^2, U^2]$ and all entries are positive.

Let $B_0 = M_X \otimes M_Y$ be the product of the conditional marginals. Undo the product tilt with $w(x,y) = 1/(\ell_x r_y) \in [U^{-2}, \delta^{-2}]$. Then $\mathcal{T}_w(M) = J_d$, and $B_1 := \mathcal{T}_w(B_0)$ is a product probability table. Lemma 16 gives

$$\|J_d - B_1\|_1 \le \frac{2U^2}{\delta^2} \|M - B_0\|_1 . \tag{79}$$

Write $B_1 = p \otimes q$. Marginalization contracts ℓ^1 distance, so both $\|p - \pi\|_1$ and $\|q - \pi\|_1$ are at most $\|J_d - B_1\|_1$. Product telescoping gives

$$\|p \otimes q - \pi \otimes \pi\|_1 \le \|p - \pi\|_1 + \|q - \pi\|_1 \le 2 \|J_d - B_1\|_1 .$$

Hence $\|J_d - \pi \otimes \pi\|_1 \le 3 \|J_d - B_1\|_1$. Combining this with (77) and (79) yields

$$\|M - M_X \otimes M_Y\|_1 \ge \frac{\delta^2 \rho^d}{6U^2 F^2}.$$

Pinsker's inequality in nats gives

$$I_M(X;Y) = D_{\mathrm{KL}}(M \| M_X \otimes M_Y) \ge \frac{1}{2} \|M - M_X \otimes M_Y\|_1^2 \ge \frac{\delta^4 \rho^{2d}}{72U^4F^4} \ge c_0 \rho^{2L}.$$

This estimate is uniform over all allowed boundary distances and values, not merely over their target average. □

D.2.9 THE FULL DETERMINISTIC ADAPTIVE CONVERSE

Fix any $\sigma \in \mathcal{D}_R$. The history identity (4) and online reduction (23) apply on any finite alphabet. Lemma 17 supplies their remaining hypothesis: every eligible same-round pair at distance at most L has conditional mutual information at least $c_0 u^2$, uniformly over the revealed boundary values. Hence

$$D_{\mathrm{KL}}(P_n \| Q_\sigma) \ge c_0 u^2 \left\lfloor \frac{n}{2L} \right\rfloor = c_0 u^2 \left\lfloor \frac{s+1}{2} \right\rfloor \ge \frac{c_0}{2} s u^2 = c_K t.$$

This covers the entire deterministic adaptive class, including value-dependent early completion. Together with (52), it proves (45) and completes the quantitative part of Theorem 12.

D.2.10 CONSEQUENCES: FIXED DEPENDENCE, PRODUCTS, AND ONE EXTRA ROUND

For fixed $t_0 > 0$ and $s = \lfloor t_0/u^2 \rfloor$, one has $t_0 - u^2 < t \le t_0$ and eventually $s \ge 1$ and $Bu \le 1/2$. Equations (46)–(47) with $\tau = t_0$ give the stated $O_{K,t_0}(L^{-1})$ bound, whereas (45) gives the positive deterministic lower limit. Each fixed phase is among the deterministic policies and has error in $[c_K t, Bt]$, while its hidden uniform mixture tends to the target.

Lemma 18 (Independent product targets)**.** *Suppose the strictly positive target path law factors as $P_X = P_Y \otimes P_Z$, with whole-site symbols $X_j = (Y_j, Z_j)$. Every R-round randomized X policy has a legal R-round randomized Y simulator with the same returned Y law.*

Proof. Write I for the original scheduling tape. Give the simulator the external tape $S = (I, U_1, \dots, U_n)$, where the U_j are independent uniforms, independent of I and all private Y-oracle coins. At an old reconstructed history (y_A, z_A), choose the original policy's batch B before reading any unused tape entry or current oracle value. Query the exact Y singleton oracle on B and internally sample each Z_j, $j \in B$, from $P_Z(Z_j \mid z_A)$ using U_j. All these conditionals use the same pre-batch history; only then is the whole batch added to the history. The fixed target's Z conditionals require internal computation, not additional Y-oracle rounds.

For equality of laws, use the filtration exposing I, completed Y outputs, and only the tape entries already used. Unused entries remain independent uniforms, so the next joint batch has law

$$\prod_{j \in B} P_Y(y_j \mid y_A) P_Z(z_j \mid z_A) = \prod_{j \in B} P_X((y_j, z_j) \mid x_A).$$

Induction gives the original joint history law after averaging the tape. This argument does not condition on the full tape. For legality, instead fix the full tape S: on every reachable Y history (A, y_A), replaying from the start uniquely reconstructs past batches and z_A. Indeed, two compatible replays cannot first diverge when their earlier batches, supplied Y values, and tape agree. Thus the next batch is a deterministic function of S and the permitted Y history. Every execution is an execution of the original policy, so reveal-once sampling and the pathwise R-round cap are preserved. Positivity ensures that each reconstructed value assignment is an admissible original execution. □

For the products in Section D.1, stationarity gives $P_{X,n} = P_{Y,n} \otimes P_{Z,n}$. If the nonconstant eigenvalues of K_Z are ζ_j, the product eigenvalues are $1, \rho, \zeta_j, \rho\zeta_j$; hence its absolute nonconstant spectral radius is ρ. Applying data processing to the returned laws in Lemma 18, the binary bounds in Appendix B.4 and (81) give, for every $\sigma \in \mathcal{R}_R^X$ and all $R, s \geq 1$,

$$D_{\mathrm{KL}}(P_{X,n} \| Q_\sigma) \geq c_Y t/L, \qquad \mathsf{h}^2(P_{X,n}, Q_\sigma) \geq 1 - e^{-b_Y t/L},$$

where $c_Y, b_Y > 0$ depend only on K_Y and $\mathsf{h}^2(P, Q) = 1 - \sum_x \sqrt{P(x)Q(x)}$. Together with Theorem 10 and $\mathsf{h}^2 \leq D_{\mathrm{KL}}/2$, these give matching randomized orders t/L for bounded t and $B_X \rho^L \leq 1/2$, where $B_X = 1/\min \pi_X = B_Y B_Z$. The simulator is randomized even if the original policy is deterministic: its revealed Z values can randomize subsequent binary scheduling. Deterministic lower bounds therefore use Theorem 10, not deterministic projection.

One additional deterministic round. Put $m = \lfloor s/2 \rfloor$. In the first round, sample positions $(2j+1)L$, $0 \leq j \leq m$, independently from π, then complete the intervals exactly. Interior gaps have $2L - 1$ unknown sites; the left tail has $L - 1$, and the right tail has zero or L sites according as s is even or odd. Since $2L - 1 = 2^R - 1$, completion takes at most R further rounds, including when $L = 1$. The stationary first-round marginal identity and lag bound in (52) give

$$D_{\mathrm{KL}}(P_n \| Q^+) = m I_{2L} \leq B m u^4 \leq \frac{B}{2} t u^2, \qquad \mathrm{TV}(P_n, Q^+) \leq \frac{\sqrt{B}}{2} \sqrt{t}\, u.$$

For $s = 1$ there is one first-round position and $Q^+ = P_n$. Thus, for every kernel in Theorem 10, an $(R+1)$-round deterministic sampler converges in both metrics as $L \to \infty$ with bounded t. The separation concerns accuracy at the same round cap.

D.3 Separation in Hellinger and total variation distance

We prove Corollary 11 by a separate affinity argument: a KL lower bound alone does not imply TV separation. Reuse the boundary-uniform pair estimate from Appendix D.2, turn it into a conditional affinity contraction, and compose contractions along the adaptive tree. The binary randomized case instead uses the marginal statistic developed in Appendices B.3 and B.4. Write

$$\mathsf{A}(P, Q) = \sum_x \sqrt{P(x)Q(x)}, \quad \mathsf{h}^2(P, Q) = 1 - \mathsf{A}(P, Q), \qquad \mathrm{TV}(P, Q) = \tfrac{1}{2} \sum_x |P(x) - Q(x)|.$$

The sampling interface and round cap remain those of Section 3.

Theorem 19 (Bounded-discrimination lower bounds)**.** *Fix a dependent, strictly positive reversible kernel K on a finite alphabet, with stationary law π, and let P_n be its stationary path law. Put $\rho = \max_{j\geq 2} |\lambda_j(K)| \in (0,1)$. For integers $R, s \geq 1$, set $L = 2^{R-1}$, $n = (s+1)L$, $u = \rho^L$, and $t = su^2$. There is a constant $a_K > 0$, depending only on K, such that*

$$\inf_{\sigma \in \mathcal{D}_R} \mathsf{h}^2(P_n, Q_\sigma) \geq 1 - e^{-a_K t}. \tag{80}$$

If, in addition, K is binary with positive nonconstant eigenvalue, there is a constant $b_K > 0$, depending only on K, such that

$$\inf_{\sigma \in \mathcal{R}_R} \mathsf{h}^2(P_n, Q_\sigma) \geq 1 - e^{-b_K t/L}. \tag{81}$$

These lower bounds require neither small u nor bounded t.

D.3.1 Uniform pair dependence and conditional path coupling

For the general finite-state kernel, define

$$d_0 = \min_{x,y} \frac{K(x,y)}{\pi(y)} > 0, \qquad U = \max_{x,y} \frac{K(x,y)}{\pi(y)}.$$

Every row has π-weighted density mean one, so $d_0 \leq 1 \leq U$. Averaging the one-step density against a probability row gives $d_0 \leq K^d(x,y)/\pi(y) \leq U$ for every integer $d \geq 1$. Choose a real eigenfunction f with eigenvalue θ, $|\theta| = \rho$, normalized by $\mathbb{E}_\pi f = 0$ and $\mathbb{E}_\pi f^2 = 1$, and let $F = \max_x |f(x)|$. Set

$$\beta = \frac{d_0^2}{12U^2F^2}, \qquad \epsilon = \frac{d_0^2}{U}, \qquad H_K = 1 + \frac{2(1-\epsilon)}{\epsilon}, \qquad a_K = \frac{\beta^2}{4(H_K^2+1)}. \tag{82}$$

Here a *pin* is a revealed coordinate together with its fixed conditioning value. For two unpinned sites at distance d with no old pin between them, let M be their conditional table and $N = M_X \otimes M_Y$. The inverse-tilting argument in Appendix D.2 gives

$$\mathrm{TV}(M,N) \geq \beta\rho^d. \tag{83}$$

Specifically, the ℓ^1 estimate in the proof of Lemma 17, with $\delta = d_0$, gives (83) after division by two. It already handles negative and repeated dominant modes.

We also need a coupling uniform in all pin assignments. Change one pin at position j and retain all others. In the adjacent interval to its right, the transition at a free site has the form

$$T(x,y) = \frac{K(x,y)r(y)}{(Kr)(x)}, \qquad r(y) = K^D(y,z)/\pi(z),$$

where z is the next fixed boundary state and $D \geq 1$ is its distance from the free site. If that boundary is absent, take $r \equiv 1$. In both cases $d_0 \leq r \leq U$, so $T(x,y) \geq \epsilon\pi(y)$. At each step, use a common Bernoulli(ϵ) coin to draw both states from the same π sample; otherwise couple the residual rows. Once the paths meet, use identical subsequent draws. Each whole path has the correct conditional bridge law, and the disagreement probability d sites from the changed pin is at most $(1-\epsilon)^d$. The fixed terminal pin is assigned identically; the minorization is not applied to a $D = 0$ step. When $\epsilon = 1$, only the common draw is needed.

For the left interval use the reversed stationary kernel, which equals K by reversibility. Conditional independence of the unpinned intervals combines these couplings with identical draws on unaffected intervals. Including the changed pin, the expected total Hamming distance is at most $1 + 2\sum_{d\geq 1}(1-\epsilon)^d = H_K$. This is a joint coupling of complete conditional paths and remains valid with any additional pins.

D.3.2 A conditional affinity contraction

Fix an old history $h = (A, x_A)$ and its selected batch B. The online charging rule of Appendix D.2 selects k pairs with disjoint endpoints, distance at most L, and no old pin between their endpoints.

For each pair let M_ℓ be its true conditional table and N_ℓ the product of its conditional marginals. Define

$$g_\ell(x,y) = \mathbf{1}\{M_\ell(x,y) \geq N_\ell(x,y)\}, \qquad S = \sum_{\ell=1}^{k} \big[g_\ell(X_{i_\ell}, X_{j_\ell}) - \mathbb{E}_{N_\ell} g_\ell\big].$$

These tests and centering constants are fixed by the old history and are not recomputed during the exposures below. With $p_B = P_n(X_B \mid h)$ and $q_B = \prod_{j\in B} P_n(X_j \mid h)$,

$$\mathbb{E}_{q_B} S = 0, \qquad \mathbb{E}_{p_B} S = \sum_{\ell=1}^{k} \mathrm{TV}(M_\ell, N_\ell) \geq k\beta u.$$

Changing one endpoint changes S by at most one, since it belongs to only one pair. Under p_B, expose only the $2k$ endpoints in a fixed order. The old history and already exposed endpoints are pins; other coordinates in B remain unconditioned. For any two values of the next endpoint, the complete-path coupling above bounds the difference of the conditional means of S by H_K. Thus each Doob increment has conditional range length at most H_K, even on a multivalued alphabet. Under q_B, endpoints are independent and the range length is at most one. Conditional Hoeffding bounds, for every real λ, give

$$\mathbb{E}_{p_B} e^{\lambda(S - \mathbb{E}_{p_B} S)} \leq e^{\lambda^2 k H_K^2/4}, \qquad \mathbb{E}_{q_B} e^{\lambda S} \leq e^{\lambda^2 k/4}.$$

No independence between the true pairs is assumed. For $\lambda \geq 0$, Cauchy–Schwarz gives

$$A(h) := \sum_{x_B} \sqrt{p_B(x_B) q_B(x_B)} \leq \sqrt{\mathbb{E}_{p_B} e^{-\lambda S} \mathbb{E}_{q_B} e^{\lambda S}} \leq e^{-\lambda k\beta u/2 + \lambda^2 k(H_K^2+1)/8}.$$

Taking $\lambda = 2\beta u/(H_K^2+1)$ yields, also for $k = 0$,

$$A(h) \leq \exp\left\{-\frac{k\beta^2 u^2}{2(H_K^2+1)}\right\}. \tag{84}$$

D.3.3 Composing contractions on the deterministic adaptive tree

By the cylinder-history argument in Appendix B.2.5, target replay factors P_n into the true batch kernels p_B, while the generated law factors into q_B on the same deterministic policy tree. Normalize the geometric mean at each node:

$$r_B(x_B \mid h) = \frac{\sqrt{p_B(x_B) q_B(x_B)}}{A(h)}.$$

Strict positivity makes these probability kernels well-defined. Let T be their path law on the same tree, padding completed paths with empty batches of affinity one. Summing full output paths gives exactly

$$\mathsf{A}(P_n, Q_\sigma) = \mathbb{E}_{\mathsf{T}} \prod_{r<R} A(h_r). \tag{85}$$

Use the online charging rule of Appendix B.2.8, with cells of length $2L$. Lemma 6 ensures that every complete path charges each cell exactly once. This is a geometric consequence of the pathwise round cap, so it also holds under T, not merely under P_n or Q_σ. Consequently,

$$\mathsf{A}(P_n, Q_\sigma) \leq \exp\left\{-\frac{\beta^2 u^2}{2(H_K^2+1)} \left\lfloor \frac{n}{2L} \right\rfloor\right\} \leq e^{-a_K t},$$

using $\lfloor n/(2L) \rfloor = \lfloor (s+1)/2 \rfloor \geq s/2$. This proves (80). The next subsection obtains a randomized marginal bound using an output statistic.

D.3.4 The binary randomized marginal bound

Now restrict to a stationary binary kernel with positive nonconstant eigenvalue. Relabel the rarer state as 1, so $\pi = (q, p)$, $q = 1 - p$, $0 < p \leq 1/2$, and $K = \rho I + (1-\rho)\mathbf{1}\pi$. Use the raw indicators

$X_i \in \{0,1\}$. The constants of Appendix B.4 are

$$m_0 = 1-\rho, \qquad M_{\max} = 1/p, \qquad \gamma = \frac{pq\, m_0^4}{M_{\max}^4},$$
$$H_{\mathrm{bin}} = 1 + 2(M_{\max}/m_0)^2 \frac{\rho}{1-\rho}, \qquad C_0 = \frac{2H_{\mathrm{bin}}}{1-\rho}.$$

Define the binary constant

$$b_K = \frac{\gamma^2}{16 C_0^2}. \tag{86}$$

It is distinct from the general finite-state constant a_K. Let

$$F_L(x) = \sum_{d=1}^{L} \rho^{L-d} \sum_{i=1}^{n-d} x_i x_{i+d}, \qquad \mu = \mathbb{E}_{P_n} F_L, \qquad \delta = \left\lfloor \frac{n}{2L} \right\rfloor \gamma u.$$

Appendix B.4 bounds each coordinate oscillation of $\mathbb{E}_{P_n}[F_L \mid \text{pins}]$ by C_0, uniformly over all pins. Exposing all n target coordinates and applying conditional Hoeffding gives

$$\mathbb{E}_{P_n} e^{-\lambda(F_L-\mu)} \le e^{\lambda^2 C_0^2 n/8}.$$

For every $\sigma \in \mathcal{R}_R$, the same appendix proves the *output-marginal* bound

$$\mathbb{E}_{Q_\sigma} e^{\lambda(F_L-\mu)} \le e^{-\lambda\delta + \lambda^2 C_0^2 n/8}, \qquad \lambda \ge 0.$$

Its scheduling tape is independent of the private singleton sampling coins; F_L never reads that tape. Applying Cauchy–Schwarz directly to the two output laws yields

$$\mathsf{A}(P_n, Q_\sigma) \le \sqrt{\mathbb{E}_{P_n} e^{-\lambda(F_L-\mu)} \mathbb{E}_{Q_\sigma} e^{\lambda(F_L-\mu)}} \le e^{-\lambda\delta/2 + \lambda^2 C_0^2 n/8}.$$

At $\lambda = 2\delta/(C_0^2 n)$ this is $\exp\{-\delta^2/(2C_0^2 n)\} \le e^{-b_K t/L}$, using $\lfloor (s+1)/2 \rfloor \ge s/2$ and $s/(s+1) \ge 1/2$. This proves (81) without a small-power restriction.

D.3.5 Matching Hellinger orders and bounded output tests

Jensen's inequality gives $D_{\mathrm{KL}}(P\|Q) \ge -2\log \mathsf{A}(P,Q) \ge 2\mathsf{h}^2(P,Q)$. The fixed-phase upper bound in Theorem 10 is Bt, where $B = 1/\min_x \pi(x)$. Hence, for every fixed $\tau > 0$ and $0 < t \le \tau$,

$$\inf_{\sigma \in \mathcal{D}_R} \mathsf{h}^2(P_n, Q_\sigma) = \Theta_{K,\tau}(t)$$

for every kernel in Theorem 19. For the positive-correlation binary subclass and the independent products of Section D.1, if also $Bu \le 1/2$, the hidden-phase KL upper bound and the randomized lower bound ((81) or Lemma 18) give

$$\inf_{\sigma \in \mathcal{R}_R} \mathsf{h}^2(P_n, Q_\sigma) = \Theta_{K,\tau}(t/L).$$

Here $B = 1/\min \pi$ is computed for the full target; in the product case, $B = B_Y B_Z$. For the lower orders use $1 - e^{-cz} \ge (1-e^{-c\tau})z/\tau$ for $0 < z \le \tau$. The deterministic order does not require the small-power condition.

Finally, $\mathrm{TV}(P,Q) \ge \mathsf{h}^2(P,Q)$ and Pinsker's inequality give, for every general finite-state kernel above,

$$\inf_{\sigma \in \mathcal{D}_R} \mathrm{TV}(P_n, Q_\sigma) \ge 1 - e^{-a_K t}, \qquad \mathrm{TV}(P_n, \overline{Q}) = O_{K,\tau}(\sqrt{t/L}), \tag{87}$$

where only the hidden-phase upper bound requires $0 < t \le \tau$ and $Bu \le 1/2$. Along $L \to \infty$ and $t \to t_0 > 0$, the fixed-kernel upper-bound conditions eventually hold for any fixed $\tau > t_0$, proving Corollary 11. The distinguishing bounded test is chosen for each deterministic policy. Together, these results give matching Hellinger orders on the binary and independent-product classes and a TV separation for reversible finite-state chains.

Interpreting one additional round. Still in the positive-correlation binary setting, write $E_R = \inf_{\sigma\in\mathcal{R}_R} \mathrm{TV}(P_n, Q_\sigma)$ and $D_{R+1} = \inf_{\sigma\in\mathcal{D}_{R+1}} \mathrm{TV}(P_n, Q_\sigma)$ for the same target P_n. The raw statistic satisfies $0 \le F_L \le n/(1-\rho)$ and $\mathbb{E}_{P_n} F_L - \mathbb{E}_{Q_\sigma} F_L \ge \delta$; the latter follows by right differentiation at zero of the generated-law MGF bound above. Since $\lfloor (s+1)/2 \rfloor/(s+1) \ge 1/3$, its normalization to $[0,1]$ and $\mathrm{TV} \ge \mathsf{h}^2$ yield, for $0 < t \le \tau$,

$$E_R \ge \frac{c}{L}\max\{u,t\}, \qquad c = \min\left\{\frac{\gamma(1-\rho)}{3}, \frac{1-e^{-b_K\tau}}{\tau}\right\} > 0.$$

Here the second term uses the concavity of $1-e^{-b_K z}$ on $0 \le z \le \tau$, with $z = t/L$. The extra-round construction in Appendix D.2 gives $D_{R+1} \le (\sqrt{B}/2)\sqrt{t}\,u$, so

$$\frac{D_{R+1}}{E_R} \le \frac{\sqrt{B}}{2c} L \min\{\sqrt{s}\,u, 1/\sqrt{s}\} \le \frac{\sqrt{B}}{2c} L\rho^{L/2} \longrightarrow 0.$$

The limit is uniform over integers $s \ge 1$ with $s\rho^{2L} \le \tau$, for fixed K, τ as $L = 2^{R-1} \to \infty$; no small-power condition is needed. When $s = 1$, $D_{R+1} = 0$ while $E_R > 0$. Thus interlacing holds for all sufficiently large L.

E Finite verification and optimality certificates

E.1 Certifying the complete two-round optimum

The supporting-hyperplane inequality for the convex mixture objective gives finite-dimensional certificates for any strictly positive finite target under the reveal-once singleton interface. Applied to the symmetric chain, it yields the complete two-round comparisons in Table 2.

Why the comparison is complete. With at most two rounds, the first set determines the entire scheduling policy. The first set B is selected before any values are seen; after its values are drawn, any remaining positions must form the second batch. For a fixed $B \ne \varnothing$, the output law is

$$Q_B(x) = \prod_{j\in B} P(X_j = x_j) \prod_{j\notin B} P(X_j = x_j \mid X_B = x_B). \tag{88}$$

The choice $B = [n]$ includes one-round termination. Thus $2^n - 1$ first-set choices enumerate all deterministic output laws, possibly with duplicates. Conditioning on any independent scheduling tape selects one such first set; consequently every hidden randomized two-round law, including arbitrary tape distributions, belongs to their convex hull. Conversely, every convex combination is implemented by privately drawing its first set with the corresponding weights. The forced complement in the second round makes this characterization specific to the two-round class.

A global optimality certificate. Let w be any simplex vector, $q_w = \sum_B w_B Q_B$, and $f(w) = D_{\mathrm{KL}}(P\|q_w)$. All laws are positive. Define

$$r_B(w) = \sum_x \frac{P(x)Q_B(x)}{q_w(x)}, \qquad g(w) = \max_B r_B(w) - 1. \tag{89}$$

Since $\partial_B f(w) = -r_B(w)$ and $\sum_B w_B r_B(w) = 1$, convexity gives, for every simplex vector v,

$$f(v) \ge f(w) + \nabla f(w)^\top (v-w) = f(w) + 1 - \sum_B v_B r_B(w) \ge f(w) - g(w).$$

Therefore

$$\max\{0, f(w) - g(w)\} \le \min_v f(v) \le f(w). \tag{90}$$

The maximization in (89) ranges over every nonempty first set, giving a bound over the full simplex.

Table 2: Certified two-round KL optima (nats).

n	ρ	D^*_{det}	ℓ_{cert}	u_{cert}	Gap bound
4	0.25	0.001954398557	0.000963398026	0.000963398027	4.650×10^{-16}
4	0.50	0.031583942402	0.015266397360	0.015266397361	9.413×10^{-15}
4	0.75	0.148551158259	0.050630805779	0.050630805780	3.595×10^{-13}
6	0.25	0.003908797113	0.001938429737	0.001938429738	1.487×10^{-15}
6	0.50	0.063167884804	0.030482742636	0.030482742638	3.391×10^{-14}
6	0.75	0.164399409572	0.121075266112	0.121075266113	4.273×10^{-13}
8	0.25	0.005863195670	0.002918193952	0.002918193953	2.035×10^{-14}
8	0.50	0.094751827206	0.046758172589	0.046758172590	2.454×10^{-13}
8	0.75	0.328798819143	0.183916975383	0.183916975388	4.415×10^{-12}

Exact certification after numerical search. Numerical optimization proposes weights for exact certification. Position reflection maps a first set B to $B^{\mathrm{rev}} = \{n+1-j : j \in B\}$. We group each first set with its reflection; each such orbit contains one or two distinct first sets. The saved weights are made exactly feasible by rounding the total weight of each orbit to a multiple of 10^{-12}, preserving total mass one, and splitting that weight equally among its first sets. For rational ρ, P, all Q_B, q_w, and every score in (89) are then rational. The score maximum and $g(w)$ can be evaluated exactly. For rational $r > 0$, logarithms admit rational enclosures from

$$\log r = 2\sum_{k=0}^{N-1} \frac{z^{2k+1}}{2k+1} + \varepsilon_N, \qquad z = \frac{r-1}{r+1}, \qquad |\varepsilon_N| \le \frac{2|z|^{2N+1}}{(2N+1)(1-z^2)}.$$

Thus a rational logarithm enclosure $a \le f(w) \le b$ produces the rigorous global interval $[\max\{0, a - g(w)\}, b]$. Separate enclosures of every deterministic law identify its minimum. The resulting interval encloses the optimum over all mixture weights.

Certified intervals. Exact enumeration of all output laws indexed by first sets and the dual bound above give Table 2. A separate implementation reconstructs singleton conditionals by summing complete target states and verifies the rational laws, weights, scores, and logarithm enclosures. Here ℓ_{cert} lower-bounds the randomized optimum and u_{cert} upper-bounds a feasible rational mixture's error. Endpoints are rounded outward; deterministic minima are displayed to 12 decimals. The last column bounds the exact dual gap before display rounding. Enumeration is exponential in n and certifies the complete two-round class, not the randomized class at larger round budgets.

A simple nonuniform KL mixture. Mixing first sets $\{2\}, \{3\}, \{1,3\}, \{2,4\}$ with weights $1/8, 1/8, 3/8, 3/8$ gives KL approximately 0.015319 at $n = 4, \rho = 1/2$, close to the certified class optimum of 0.015266. This improves on the uniform phase mixture without solving the full convex program.

E.1.1 An exact four-site total-variation separation

Fix $n = 4$, $\rho = 1/2$, and a pathwise round cap $R = 2$ for the stationary symmetric binary chain with $K(x,y) = (1+\rho xy)/2$, $x, y \in \{-1, 1\}$. Write $\mathrm{TV}(P,Q) = \frac{1}{2}\sum_x |P(x) - Q(x)|$. Every deterministic policy is specified by its nonempty first set $A \subseteq \{1,2,3,4\}$: its last round must fill the complement. The case $A = \{1,2,3,4\}$ includes one-round sampling. Thus the complete hidden-randomized class is the simplex over the following 15 laws:

$$Q_A(x) = 2^{-|A|} \prod_{j \notin A} P(X_j = x_j \mid X_A = x_A).$$

Exact rational enumeration and a rational primal–dual certificate give

Sampler or policy class	Total variation
Best deterministic policy	$1/8 = 0.125$
Uniform grid $Q_{\mathrm{grid}} = (Q_{\{1,3\}} + Q_{\{2,4\}})/2$	$7/80 = 0.0875$
Best hidden-randomized policy	$1977/26720 \approx 0.073990$

The deterministic minimum is attained exactly by first sets $\{1,3\}$ and $\{2,4\}$. In particular, at the concrete tolerance $\mathrm{TV} < 0.1$, the original uniform-grid construction already qualifies, whereas no deterministic policy under the same cap does.

An exact optimizer over all hidden mixtures is

$$Q_* = \frac{1552(Q_{\{2\}} + Q_{\{3\}}) + 465(Q_{\{1,3\}} + Q_{\{2,4\}}) + 2646Q_{\{1,4\}}}{6680}.$$

For a dual certificate, define a bounded test $f : \{-1,1\}^4 \to [0,1]$ by its sign strings:

$$f(x) = \begin{cases} 1, & x \in \{----, ++++\}, \\ 343/2505, & x \in \{--++, ++--\}, \\ 212/835, & x \in \{-++-, +--+\}, \\ 0, & \text{otherwise.} \end{cases}$$

This test is invariant under global spin flip and spatial reversal. Exact evaluation gives

$$\mathbb{E}_P f = \frac{2421}{5344}, \qquad \max_{\varnothing \neq A \subseteq \{1,2,3,4\}} \mathbb{E}_{Q_A} f = \frac{633}{1670}.$$

Consequently every hidden mixture Q satisfies

$$\mathrm{TV}(P,Q) \geq \mathbb{E}_P f - \mathbb{E}_Q f \geq \frac{2421}{5344} - \frac{633}{1670} = \frac{1977}{26720}.$$

The law Q_* attains this bound. All 15 component laws, deterministic TV values, weights, and dual inequalities are checked using rational arithmetic; floating-point linear programming is used only to discover the certificate. The artifact's standalone checker reconstructs all component laws from nearest-pin conditionals and verifies the certificate independently of the optimization code.

E.2 Computing the full deterministic optimum

The two-round certificates enumerate the first set. For more rounds, a policy can adapt later batches to the values it has generated. This appendix computes the optimum over that entire deterministic class without enumerating its decision trees. Boundary-conditioned Bellman planning on chains is classical (Krause & Guestrin, 2005), and Wen et al. (2026) give an exact adaptive forward-KL dynamic program for uniform permutations. Here the Markov reduction combines a shared parallel horizon with the total-correlation cost of an arbitrary simultaneous batch.

Proposition 20 (Finite-alphabet interval optimization)**.** *Let $P_n(x_{1:n}) = \pi(x_1)\prod_{i=1}^{n-1} K(x_i, x_{i+1})$ be a stationary homogeneous Markov law on an alphabet $\mathcal{X}$ of size q, with $K(x,y) > 0$ and $\pi K = \pi$. The minimum of $D_{\mathrm{KL}}(P_n \| Q_\sigma)$ over $\sigma \in \mathcal{D}_R$ is the value of an interval dynamic program with $O(R(q+1)^2 n)$ scalar states and $O(nq^3 + Rn^3(q+1)^2 q^2)$ arithmetic operations, treating logarithm evaluation as a scalar operation. Each simultaneous action is obtained by a shortest path through the interval. An optimal policy can be chosen whose decisions within each gap depend only on its length, its revealed boundary values, and the common remaining round budget. Reversibility is not required.*

The resource is a pathwise round cap: all gaps may act in the same round, with no additional constraint on batch width. The operation bound concerns the recurrence, not the bit complexity of certified logarithms or the cost of enumerating its full output law.

E.2.1 Target replay and interval separation

For a history $h = (A, x_A)$, let $W_k(h)$ be the smallest expected remaining conditional-total-correlation cost when a target assignment is exposed according to a deterministic policy with at most k rounds left. Equation (4) identifies its initial value with the desired forward KL. Its Bellman equation is

$$W_k(h) = \min_{\varnothing \neq B \subseteq [n] \setminus A} \left\{ \mathrm{TC}_P(X_B \mid h) + \mathbb{E}_{P(X_B|h)} W_{k-1}(A \cup B, x_A, X_B) \right\}. \tag{91}$$

Completed histories have value zero; an incomplete history at $k = 0$ has value $+\infty$. The transition expectation is under the target joint conditional, even though the generator draws the selected sites

independently. This distinction follows directly from the forward-KL identity and is essential to the objective.

Write $V_k(\ell, a, b)$ for the local value in a gap of ℓ sites with boundary values $a, b \in \mathcal{X} \cup \{\varnothing\}$, where $\varnothing$ denotes a physical end of the chain. Empty gaps have value zero and nonempty gaps at $k = 0$ have value $+\infty$. Conditional on h, the remaining gaps are independent Markov bridges, so

$$W_k(h) = \sum_{J \text{ a gap of } A} V_k(|J|, a_J, b_J). \tag{92}$$

To prove this equality, induct on k. Split any chosen batch among the current gaps. Conditional total correlation adds across the gaps, as does the inductive continuation value after the reveal. Its target expectation therefore separates into one cost for each local subset. A choice informed by other gaps' earlier values is still one such subset and cannot improve on the local minimum. A gap left untouched contributes V_{k-1}, which is at least V_k because the horizon is an upper bound. Conversely, take a minimizing first action in every nonempty gap and execute their union, then run their optimal continuations in parallel. Initial idle rounds can be removed from each local policy. This attains the sum with the same k global rounds, proving both directions. In particular, the reduction does not assume away cross-gap adaptation.

E.2.2 A SHORTEST PATH CHOOSES THE WHOLE NEXT BATCH

Fix a parent gap (ℓ, a, b) and let ν be its target conditional law. For ordered selected positions $i_1 < \cdots < i_m$, the marginal $(X_{i_1}, \ldots, X_{i_m})$ is a Markov chain. Its entropy factorization gives

$$\mathrm{TC}_\nu(X_{i_1}, \ldots, X_{i_m}) = \sum_{j=1}^{m-1} I_\nu(X_{i_j}; X_{i_{j+1}}).$$

The mutual informations use the parent conditional law, before any of the new values have been observed. Each child gap depends only on its two endpoints. Thus define, for $i < j$,

$$\begin{aligned} c_{\mathrm{start}}(i) &= \sum_x \nu_i(x) V_{k-1}(i-1, a, x), \\ c(i,j) &= I_\nu(X_i; X_j) + \sum_{x,y} \nu_{ij}(x,y) V_{k-1}(j-i-1, x, y), \\ c_{\mathrm{end}}(i) &= \sum_x \nu_i(x) V_{k-1}(\ell - i, x, b). \end{aligned}$$

The local objective for a selected subset is exactly its start cost, the sum of consecutive edge costs, and its end cost. Set

$$d(j) = \min\left\{ c_{\mathrm{start}}(j), \min_{i<j}[d(i) + c(i,j)] \right\}, \qquad V_k(\ell, a, b) = \min_j [d(j) + c_{\mathrm{end}}(j)].$$

A shortest-path backtrace chooses the entire simultaneous batch; no selected value is consulted while constructing it. At $k = 1$, finite continuation cost forces the path to visit every remaining position. There are at most $(q+1)^2$ boundary types for each length and horizon. A gap of length ℓ has $O(\ell^2)$ candidate edges; each continuation expectation and mutual information requires $O(q^2)$ scalar work. Summing over lengths and horizons gives $O(Rn^3(q+1)^2q^2)$ operations, in addition to $O(nq^3)$ for transition powers. Parent marginals and mutual informations can be shared across horizons. Scalar values require $O(Rn(q+1)^2)$ storage; explicitly storing all selected batches can require $O(Rn^2(q+1)^2)$. Transition powers use $O(nq^2)$ storage, and materializing one parent's pair marginals uses $O(n^2q^2)$ workspace. These are arithmetic counts, not the bit complexity of rational powers, certified logarithms, or full-law enumeration.

For a gap with positions $1, \ldots, \ell$, put

$$L_i(x) = \begin{cases} K^i(a,x), & a \neq \varnothing, \\ \pi(x), & a = \varnothing, \end{cases} \qquad T_i(x) = \begin{cases} K^{\ell+1-i}(x,b), & b \neq \varnothing, \\ 1, & b = \varnothing. \end{cases}$$

The common normalizer is $Z = K^{\ell+1}(a,b)$ when both boundaries are present, $Z = \pi(b)$ when only the right boundary is present, and $Z = 1$ when the right boundary is absent. Hence the parent

marginals are

$$\nu_i(x) = \frac{L_i(x)T_i(x)}{Z}, \qquad \nu_{ij}(x,y) = \frac{L_i(x)K^{j-i}(x,y)T_j(y)}{Z}.$$

In particular, an absent left boundary contributes π, not a uniform law. The existing shortest-path costs apply with sums over $\mathcal{X}$ and $\mathcal{X}^2$.

For the symmetric binary kernel (2), the bridge marginals simplify further. For a single position i, let $\alpha = a\rho^i$ if the left endpoint is present and zero otherwise, and $\beta = b\rho^{\ell+1-i}$ if the right endpoint is present and zero otherwise. Then

$$\nu_i(x) = \frac{(1+\alpha x)(1+\beta x)}{2(1+\alpha\beta)}.$$

For $i < j$, instead use $\alpha = a\rho^i$, $\gamma = \rho^{j-i}$, and $\beta = b\rho^{\ell+1-j}$, with the same absent-endpoint convention. Then

$$\nu_{ij}(x,y) = \frac{(1+\alpha x)(1+\gamma xy)(1+\beta y)}{4(1+\alpha\beta\gamma)}.$$

These give exact rational probabilities whenever ρ is rational.

E.2.3 Certified three-round comparisons

We apply the recurrence at $R = 3$, $n \in \{8, 12\}$, and $\rho \in \{1/4, 1/2, 3/4\}$, covering all six parameter combinations. Table 3 compares the complete deterministic optimum with the legal uniform four-phase mixture, writing $D^*_{\mathrm{det}} = \inf_{\sigma\in\mathcal{D}_R} D_{\mathrm{KL}}(P_n\|Q_\sigma)$. Each mixture upper endpoint is strictly below the corresponding deterministic lower endpoint. Displayed values are rounded from outward interval enclosures. The ratio is the deterministic optimum divided by the mixture error.

Table 3: Certified three-round KL comparisons (nats).

n	ρ	D^*_{det}	$D_{\mathrm{KL}}(P_n\|\overline{Q})$	Ratio
8	0.25	0.0000076294	0.0000020938	3.644
8	0.50	0.0019543986	0.0007042135	2.775
8	0.75	0.0509270340	0.0276655821	1.841
12	0.25	0.0000152588	0.0000042553	3.586
12	0.50	0.0039087971	0.0015058865	2.596
12	0.75	0.0819160987	0.0629505481	1.301

To enclose the costs, we evaluate rational logarithms with power-of-two range reduction and 64 terms of

$$\log u = 2\sum_{j=0}^{63} \frac{z^{2j+1}}{2j+1} + r, \qquad z = \frac{u-1}{u+1}, \quad 1 \le u \le 2, \quad 0 \le r \le \frac{2z^{129}}{129(1-z^2)}.$$

Logarithm and cost enclosures use outward-rounded integers at scale 10^{50}, with exact rational target probabilities. For candidate intervals $[l_i, u_i]$, the interval $[\min_i l_i, \min_i u_i]$ encloses their minimum. Positive target probabilities preserve these enclosures throughout the recurrence. Selecting a candidate with smallest upper endpoint produces a feasible policy within the certified tolerance of optimality; overlapping intervals are retained rather than interpreted as unique exact minimizers.

For each case, enumeration of all 2^n outputs reconstructs the selected policy's rational law, verifies normalization and agreement with the Bellman KL enclosure, and evaluates the hidden mixture. This exponential check is not needed by the polynomial recurrence. An independent uncompressed full-history implementation at $n = 8, R = 3, \rho = 1/2$ checks every visited-state cost enclosure against the interval decomposition. Direct subset and joint-entropy calculations also check the shortest-path and mutual-information identities on small bridges.

At $n = 12, \rho = 3/4$, the selected policy starts with $\{4, 9\}$, beating the grid with three first-round positions but still losing to the mixture. The selected policies choose their batches independently of the revealed boundary values.

E.3 A RATIONAL CERTIFICATE FOR THE ANALYTIC BOUNDS

The finite policy-class certificates above are complemented by a direct rational evaluation of the analytic bounds. Take $\rho = 3/10$, $L = 256$, $R = 9$, and $s = \lfloor (1/20)\rho^{-512} \rfloor$. Outward rational bounds in (11) and (12) certify

$$D_{\mathrm{KL}}(P_n \| \overline{Q}) \le 0.00449507 \quad < \quad 0.00770537 \le \inf_{\sigma \in \mathcal{D}_9} D_{\mathrm{KL}}(P_n \| Q_\sigma). \tag{93}$$

Here $n = (s+1)L \approx 6.6241 \times 10^{268}$: this is a finite witness for the analytic theorem, not a claim of practical sequence length. A tolerance of 0.006 nats is feasible for the randomized policy but not for any deterministic policy at the same depth.

All parameters entering the bounds are rational or integer. Since $t \le t_0 = 1/20$, monotonicity allows replacing t by t_0 in the upper bound. For $0 \le x < m + 2$, a positive Taylor polynomial with a geometric remainder gives

$$e^x \le \sum_{k=0}^{m} \frac{x^k}{k!} + \frac{x^{m+1}}{(m+1)!} \frac{1}{1 - x/(m+2)}.$$

A rational upper bound on each square root is certified by squaring. These operations, followed by outward rounding, prove the strict inequality using exact arithmetic. The artifact retains the complete integer parameters and rational enclosures; implementation settings, diagnostic plots, and execution instructions are documented separately from the mathematical certificate.